

El sistema geoelectoral mexicano desde la perspectiva de los sistemas complejos

The Mexican Geo-Electoral System from a Complex Systems Perspective

M. David Álvarez Hernández*
Miguel Álvarez Texcotitla**

Resumen

Desde la perspectiva de sistemas complejos, en la presente investigación se analiza el sistema geoelectoral mexicano, enfocándose en los factores que influyen en los procesos de integración, mantenimiento y actualización del Registro Federal de Electores. El análisis ofrece una definición del concepto de complejidad electoral y, a partir de la delimitación de este concepto, se construye un conjunto de índices, los cuales permiten cuantificar la complejidad electoral a partir de información procedente de diversos indicadores geoespaciales, demográficos y estructurales. Los resultados obtenidos con estos índices se complementan con la elaboración de una tipología de complejidad electoral, la cual permite clasificar las

Artículo recibido el 2 de junio de 2020 y aceptado para su publicación el 2 de julio de 2020. La **dictaminación** de este trabajo fue realizada por evaluadores externos al Instituto Electoral del Estado de México.

■ pp. 213-258



* Maestrando en Complex Systems por el King's College London. Jefe del Departamento de Evaluación Demográfica 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

** 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

distintas unidades territoriales del sistema geoelectoral mexicano en función de su complejidad electoral.

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

Abstract

This paper analyses the Mexican geo-electoral system from the perspective of complex systems, focusing on factors that influence the integration, maintenance and updating of the Federal Register of Voters. This analysis offers a definition of the concept of electoral complexity, and from the definition of this concept a set of indices is constructed, which allow to quantify the electoral complexity from information coming from diverse geospatial, demographic and structural indicators. The results obtained with these indices are complemented with the elaboration of an Electoral Complexity Typology, which allows the classification of the different territorial units that conform the Mexican geo-electoral system according to their degree of electoral complexity.

Keywords: complex systems, geo-electoral systems, socioeconomic indicators, principal component analysis, geostatistics.

Introducción

La integración, el mantenimiento y la actualización de los registros de electores son tareas fundamentales para garantizar el adecuado funcionamiento de los procesos electorales. En México, el Registro Federal de Electores (RFE) ha contribuido significativamente a otorgar mayor certeza a los procesos político-electorales que se desarrollan en el país. Asimismo, el RFE ha coadyuvado a generar una mayor legitimidad a los resultados de las elecciones, ya que uno de sus propósitos es garantizar el libre acceso y ejercicio de los derechos políticos.

En este contexto resulta primordial contar con herramientas analíticas que permitan planear, dirigir y evaluar los programas y acciones institu-

cionales enfocados a la integración, el mantenimiento y la supervisión del RFE, y que además tomen en consideración la diversidad geográfica del país y la pluralidad de las características sociodemográficas de la población.¹ Particularmente, resulta importante desarrollar indicadores que sinteticen y cuantifiquen los factores que inciden en la complejidad electoral de las secciones electorales, ya que dichas unidades territoriales son los bloques básicos que conforman al sistema geoelectoral mexicano.²

Por consiguiente, el objetivo principal del presente trabajo de investigación consiste en definir y cuantificar la complejidad electoral asociada a las características geoespaciales, demográficas y estructurales que presentan cada una de las unidades territoriales del sistema geoelectoral mexicano. Para alcanzar este objetivo se elabora un conjunto de índices de complejidad electoral seccionales (ICES), los cuales se sustentan en información procedente de diversos indicadores geoespaciales, demográficos y estructurales. Los resultados estadísticos obtenidos a partir de los ICES se complementan con la elaboración de una tipología de complejidad electoral seccional (TCES), la cual permite tipificar las secciones electorales en función de su grado de complejidad electoral.

Es importante mencionar las contribuciones principales del trabajo para dimensionar su importancia. La primera contribución es metodológica, ya que presenta una metodología unificadora inédita que permite conocer con precisión la dinámica de los factores demográficos, geoespaciales y estructurales que determinan la complejidad de las secciones electorales, facultando el análisis y el monitoreo de éstas para propósitos de estudios institucionales y académicos.

Otra contribución es que los resultados fundamentan la propuesta de utilizar la información cartográfica, demográfica y electoral que genera continuamente el INE a nivel seccional; información que se caracteriza por tener un

¹ Por mandato constitucional (Constitución Política de los Estados Unidos Mexicanos [CPEUM], artículo 41), el Instituto Nacional Electoral (INE) es el organismo público encargado de la integración, el mantenimiento y la actualización del RFE.

² El concepto de *sistema geoelectoral* hace referencia a dos cuestiones: las características de geografía humana propias de cada unidad territorial del país y las interrelaciones existentes entre dichas unidades territoriales. En particular, las secciones electorales se pueden considerar como las unidades más básicas del sistema geoelectoral (Ley General de Instituciones y Procedimientos Electorales [Legipe], artículo 147).

mejor nivel de actualización respecto a otras fuentes de información oficial. Asimismo, los resultados coadyuvan a robustecer la calidad del registro de electores de México, ya que proporcionan una forma efectiva para analizar y monitorear la evolución de los factores demográficos, geoespaciales y estructurales que afectan las tareas de integración, mantenimiento y actualización del RFE.

Por otro lado, la utilización de los ICES y la TCES permite evaluar la situación demográfica de México, ya que estos instrumentos posibilitan el análisis de la variabilidad del comportamiento poblacional de las secciones electorales del país, lo cual a su vez facilita la creación y la aplicación de modelos de pronóstico demográfico locales, que proporcionan estimaciones y proyecciones más precisas. Asimismo, los ICES y la TCES ayudan a solventar las limitaciones de los estudios que tienen como propósito monitorear la evolución del sistema geoelectoral mexicano a escala seccional.

Por lo tanto, los resultados se alcanzaron siguiendo una estrategia de investigación que inicia con una revisión del debate contemporáneo sobre los sistemas complejos, dado que la metodología, utilizada para la construcción de los ICES y la TCES, tiene sus bases en las ciencias de la complejidad y en los métodos de análisis multivariante de la ciencia de datos.

En la segunda sección se presenta el marco metodológico. Se propone una definición heurística de complejidad electoral que justifica el tipo de indicadores estadísticos seleccionados para cuantificar la complejidad del sistema geoelectoral. Asimismo, se expone la metodología de elaboración de los ICES —fundamentada en el análisis de componentes principales—, y se explica el algoritmo de agrupamiento —K-medias— utilizado en la conformación de la TCES.

En la tercera sección se ofrecen los resultados del proceso de construcción de los ICES: del análisis exploratorio de datos realizado para los indicadores base seleccionados; de la aplicación de la técnica de análisis de componentes principales, detallando la capacidad explicativa de los ICES; y, de los patrones de agrupamiento discernidos en la proyección de las secciones electorales en el espacio de los ICES.

En la cuarta sección, se presentan los resultados de la implementación del algoritmo de agrupamiento K-medias para la construcción de la TCES. En particular, se muestra el proceso de determinación de la configuración

óptima de grupos en la tipología, y se analizan las estadísticas de tendencia central de cada grupo de la tipología para identificar sus rasgos característicos y los casos atípicos de complejidad electoral.

Por último, se presentan las conclusiones que retoman, desde una perspectiva global y unificadora, 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 ICES y la TCES.

El debate sobre los sistemas complejos

El tema de la complejidad es controversial ya que, para algunos, las investigaciones sobre la complejidad representan el posible inicio de una revolución científica (por ejemplo, Waldrop, 1992; Wolfram, 2002); mientras que, para otros, la complejidad representa un paradigma sin bases científicas sólidas (por ejemplo, Horgan, 2015).

En este marco de divergencia se pueden distinguir dos corrientes (Phelan, 2001, p. 132; Becerra, 2020, p. 2): la tradición sistémico-cibernética enfocada al estudio de fenómenos de varios dominios desde una perspectiva de sistemas; y las autonombradas *ciencias de la complejidad*, caracterizadas por el enfoque algorítmico para analizar el comportamiento de sistemas adaptativos, dinámicos, no equilibrados y no lineales. Ambas corrientes no tienen un sentido unívoco en torno al concepto de *complejidad*, ni incluso al interior de ellas (Lloyd, 2001, p. 7).

Por otra parte, desde la perspectiva de las ciencias de la complejidad se ha considerado que la forma adecuada de abordar la complejidad es encontrando una definición formal que especifique las condiciones necesarias y suficientes que permitan distinguir los sistemas que son complejos de aquellos que no lo son. Alternativamente, la tradición sistémico-cibernética señala que se puede definir el concepto de complejidad a través de un enfoque más fenomenológico, el cual busque caracterizar al conjunto de sistemas complejos en términos de sus elementos constitutivos (véase 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. Por ejemplo, en las ciencias de la computación se define un concepto de *complejidad* que sólo puede ser utilizado adecuadamente en el ámbito de su disciplina (Flores Valdés y Martínez Mekler, 2011).

En relación con una definición más amplia, la cuestión principal reside en discriminar, en forma no arbitraria, cuáles sistemas 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*.

Considerando 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 se consideran como inherentes a los sistemas complejos, y que no son exclusivas de un campo de estudio específico, ya que trascienden las barreras disciplinarias y establecen conexiones entre multitud de fenómenos y sistemas de muy variada naturaleza (véanse Ladyman *et al.*, 2013; Mitchell, 2009).

El conjunto de propiedades son las siguientes:

1. La composición jerárquica de elementos constitutivos. Los sistemas complejos están constituidos por un número vasto de bloques individuales, que interactúan en formas complejas, dando lugar a la formación de estructuras jerárquicas a diferentes escalas dentro del sistema.
2. El procesamiento de información y respuesta al entorno. Los mecanismos internos de los sistemas complejos son capaces de captar y procesar la información procedente del entorno, y cambiar su comportamiento en función de los estímulos externos recibidos.
3. 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 propicias para el surgimiento de comportamientos colectivos entre los bloques constitutivos del sistema.

4. La adaptación, la resiliencia y la evolución. Los sistemas complejos presentan dinámicas internas que, en conjunto, conservan el estado actual del sistema frente a cambios exógenos.

Por consiguiente, con base en las anteriores propiedades, Mitchell (2009) propone una 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. (p. 13)

Por otra parte, a medida que los estudios sobre complejidad se han ido incrementado, el número de formas de medir la complejidad ha crecido paralelamente. Lloyd (2001) clasificó 40 medidas de complejidad, agrupadas en tres cuestiones principales: 1) dificultad de descripción, 2) dificultad de creación y 3) grado de organización.

La primera cuestión señala que la complejidad de un sistema 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. Y la tercera está vinculada con aspectos organizativos, por ejemplo, la cantidad de información compartida entre las partes de un sistema como resultado de su estructura organizativa.

La complejidad en las ciencias sociales

Los sistemas sociales son complejos considerando que una de sus características centrales es la interacción entre personas y grupos. 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. Por ende, la tarea de la investigación social es hallar estos 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.

Actualmente, en las ciencias sociales se tienen dos enfoques de investigación principales, el cuantitativo y el cualitativo, que se caracterizan por tener diferentes fortalezas y debilidades (Haynes, 2018, p. 2).

Generalmente, los enfoques cuantitativos son mejores cuando los problemas están bien delimitados, y cuando hay oportunidad de aplicar técnicas de modelado matemático/estadístico que permitan obtener resultados generales de los conceptos que se están problematizando. Por su parte, los métodos cualitativos son más apropiados cuando las preguntas iniciales son poco claras, o la definición de las preguntas y los problemas de interés exigen más debate.

Este dualismo coexiste con otros enfoques alternativos que argumentan la necesidad de crear una síntesis en la interfaz de los enfoques cuantitativo y cualitativo que tome como referencia las ideas provenientes del campo de las ciencias de la complejidad. El argumento principal de estos enfoques alternativos es que la gran mayoría de los sistemas sociales muestran las principales propiedades que caracterizan a los sistemas complejos; es decir, los fenómenos evolucionan de manera emergente y dinámica y con frecuencia no son reducibles a explicaciones causales mecanicistas que se mantengan por largos periodos (Haynes, 2018, pp. 5-33).

Por lo tanto, considerando los enfoques descritos, lo que se necesita en el ámbito de las ciencias sociales son técnicas y metodologías adaptativas, que sean capaces de ajustarse y combinarse con el pragmatismo de las ciencias de la complejidad, cuando las circunstancias lo ameriten. Técnicas y metodologías con 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.

Por otra parte, es importante mencionar que los sistemas de indicadores socioeconómicos con frecuencia son la base fundamental para cualquier descripción sólida de los hechos sociales y para la formulación de políticas.³

³ Los indicadores socioeconómicos son aquellos indicadores (métricas, índices, *rankings*, clasificaciones) elaborados con el fin de cuantificar el comportamiento global o parcial de un cierto fenómeno o sistema social/humano.

Sin embargo, la forma tradicional de construcción de estos indicadores ha estado influenciada por el paradigma reduccionista, por lo que muchos métodos de cuantificación tienden a generar medidas agregadas de los fenómenos, con el fin de simplificar el análisis.

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 agregadas 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 (Maggino, 2017, pp. VII-VIII).

La descripción y cuantificación de los fenómenos sociales debe preservar de alguna manera la complejidad inherente de estos fenómenos. 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 hacedores de políticas.

Marco metodológico

No obstante que el concepto de *complejidad* se ha exportado y utilizado en diversos campos de las ciencias sociales, su uso en temas político-electorales ha sido mínimo o se ha seguido un enfoque estrictamente computacional.⁴ Considerando esas limitaciones y la necesidad de superarlas, la presente investigación analiza desde la perspectiva de los sistemas complejos un aspecto particular, pero de suma importancia, de los sistemas geoelectorales: los registros de electores.

⁴ Faliszewski *et al.* (2006) realizan una exposición de los estudios sobre complejidad en los sistemas electorales, desde una perspectiva computacional y teórica.

En el sistema geoelectoral mexicano el RFE está conformado por dos instrumentos: el Padrón Electoral (PE) y la Lista Nominal de Electores (LNE). Para integrar, mantener y actualizar estos instrumentos es necesario efectuar periódicamente 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.

La realización de estas actividades implica considerar todos aquellos factores que obstaculizan las labores de integración, mantenimiento y actualización del RFE. Es decir, el INE debe considerar y cuantificar la complejidad que se encuentra asociada a las características geoespaciales, demográficas y estructurales que presentan cada una de las unidades territoriales del sistema geoelectoral.⁵

Por consiguiente, considerando lo anterior, se define el concepto de *complejidad electoral* como el conjunto de factores geoespaciales, demográficos y estructurales que obstaculizan la actualización, el mantenimiento y la supervisión del RFE. Un concepto de complejidad electoral que incorpora las características geoespaciales, la dinámica demográfica y el nivel de actualización de los instrumentos registrales.⁶

Una vez definida la complejidad electoral, surge la problemática de cómo cuantificar la complejidad de las secciones electorales de México. Para solucionar este problema se propone identificar, en primera instancia, los indicadores esenciales que describen la dinámica de los factores que determinan la complejidad electoral, pues a partir de estos indicadores se podrá construir un conjunto de índices/métricas que cuantifiquen el grado de complejidad electoral observado en las secciones electorales. Asimismo, estos índices/métricas servirán posteriormente como insumos principales para la elaboración de una tipificación de las secciones electorales.

⁵ Es importante mencionar que el énfasis hacia los factores geoespaciales, demográficos y estructurales responde en parte a las limitaciones de información. En consecuencia, la presente investigación no considera otros aspectos importantes del sistema geoelectoral mexicano (por ejemplo, los mecanismos de votación), debido a la falta de indicadores estadísticos que permitan incorporarlos en el concepto de *complejidad electoral*. Asimismo, se dejan de lado factores coyunturales que impactan directamente en las tareas de conformación del RFE (por ejemplo, la pandemia covid-19).

⁶ De acuerdo con la clasificación de Lloyd (2001), la definición de *complejidad electoral* que se propone cae dentro de la categoría de dificultad de descripción, ya que su cuantificación depende de los factores esenciales que describen las características del sistema geoelectoral.

Los indicadores base para cuantificar la complejidad electoral

De acuerdo con la definición heurística de *complejidad electoral* es posible distinguir tres categorías de indicadores (geoespaciales, demográficos y estructurales) que el INE monitorea constantemente para la integración, el mantenimiento y la actualización del RFE.

Indicadores geoespaciales. Estos indicadores describen primordialmente factores como el volumen y la distribución 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 las siguientes tareas institucionales: instalación de casillas electorales, programación de los operativos de campo para la actualización de las bases cartográficas y el levantamiento de estudios muestrales.

Indicadores demográficos. Éstos se encuentran asociados esencialmente a factores de la dinámica demográfica de las secciones electorales; particularmente, a los cambios demográficos observados en la LNE. Los anteriores factores son elementos fundamentales para determinar el grado de complejidad electoral de las secciones electorales, principalmente porque los programas de integración, identificación y actualización del RFE (por ejemplo, las campañas de credencialización, la identificación de flujos migratorios anómalos y los programas de reseccionamiento⁷) deben adaptarse adecuadamente a las distintas dinámicas demográficas que se observan en el sistema goelectoral.

Indicadores estructurales. Éstos se encuentran asociados primordialmente a las actividades institucionales de integración y actualización del RFE. Es importante señalar que las unidades territoriales que presenten niveles deficientes en estos indicadores requerirán una mayor asignación de recursos institucionales, esto con el fin de solventar dichas deficiencias y garantizar la consistencia de los instrumentos registrales para los procesos electorales.

⁷ Reseccionamiento se refiere al proceso de modificar los límites geográficos de las secciones electorales, con el fin de cumplir la normatividad de tener un máximo de 3000 ciudadanos por sección (Legipe, artículo 147).

A partir de estas tres categorías de indicadores, se identifica un conjunto de siete indicadores base que aportan información crucial para la cuantificación de la complejidad electoral. Es decir, estos indicadores base, que forman el conjunto de datos,⁸ permiten construir el conjunto de índices/métricas que cuantifican la complejidad electoral.⁹

La Magnitud de la LNE (LNE). Este indicador es importante ya que las actividades relacionadas con la integración y actualización del RFE dependen de las estimaciones demográficas para la elaboración de planes presupuestales. En consecuencia, las secciones electorales que superan los límites de población establecidos presentan un mayor grado de complejidad electoral, ya que dichas secciones deben ser monitoreadas periódicamente para evaluar la aplicación de estrategias de reseccionamiento.

La Densidad de la LNE (Densidad_LNE). Este indicador geoespacial y demográfico se asocia con el grado de dispersión de la población dentro de una sección electoral. Este indicador incide en el grado de complejidad electoral que debe afrontar el INE para el cumplimiento de sus actividades, ya que las secciones electorales localizadas en zonas urbanas presentan ventajas logísticas para realizar las tareas de integración y actualización del RFE, en comparación con las secciones electorales ubicadas en zonas rurales.

La razón de actualización LNE/PE (Razon_LNE_PE). Este indicador da cuenta del nivel de actualización de la LNE y el PE,¹⁰ y ayuda a explicar el nivel de complejidad electoral asociado a la implementación de estrategias institucionales que promuevan la actualización de los datos personales de los ciudadanos residentes en las secciones electorales.

La Tasa de Crecimiento Anual de la LNE (TC_LNE_2019). Existe una mayor complejidad electoral en las secciones electorales de rápido creci-

⁸ Es necesario advertir que, de un universo total de 68 436 secciones electorales existentes en 2019, sólo se pudo recuperar la información completa de los siete indicadores base para un subuniverso de 61 004 secciones electorales.

⁹ En el anexo I se encuentra la descripción matemática explícita de los siete indicadores base. Asimismo, los datos fuente de los indicadores se pueden consultar en el repositorio de GitHub (https://github.com/MDAlvarezH/Articulo_Codigo-Analisis-sistema-geo-electoral-mexicano-) y en el archivo Python_Codigo_Complejidad sistema geo-electoral_Secc_Procesamiento se explican los detalles técnicos del procesamiento de los indicadores.

¹⁰ El nivel de actualización se define como la razón entre el número de ciudadanos inscritos en la LNE y el número de ciudadanos registrados en el PE (valores entre 0 y 1). Entre menor sea este cociente, mayor será el nivel de desactualización entre la LNE y el PE.

miento demográfico, ya que dicho fenómeno impacta en los procesos de actualización de la LNE y el PE. Este indicador puede reflejar el impacto de otros fenómenos sociales, como es el caso de la migración forzada por diversas circunstancias, fenómenos que contribuyen a generar una mayor complejidad electoral.

El coeficiente de Variación de la LNE (Coef_Var_LNE). Este indicador proporciona información acerca de la variación demográfica en las secciones electorales. Se tiene un mayor grado de complejidad electoral en las secciones sujetas a un mayor nivel de variación poblacional, lo cual inherentemente afecta la planeación institucional referente a los procesos de actualización cartográfica y actualización del PE y la LNE. Las unidades territoriales que presentan una mayor variación poblacional son más complejas desde el punto de vista electoral, ya que es más difícil obtener estimaciones y pronósticos demográficos certeros, para efectos de la planeación institucional.

El coeficiente de Variabilidad Proporcional de la LNE (Var_Prop_LNE). Este coeficiente complementa la información proporcionada por el Coeficiente de Variación (cv), ya que, en lugar de basarse en la desviación de un valor promedio, la Variabilidad Proporcional (vp) se cuantifica mediante la comparación de los datos observacionales, sin la necesidad de requerir suposiciones sobre la tendencia central o las distribuciones estadísticas subyacentes.¹¹ Por lo tanto, se asume que entre mayor sea la vp de la LNE de una sección, mayor será la complejidad electoral de dicha unidad territorial.

La autocorrelación de la LNE (Autocorr_LNE). Para los propósitos de cuantificar la complejidad electoral, es de particular interés calcular la autocorrelación de la LNE con un rezago mensual (1 retraso), ya que los pronósticos de LNE se realizan frecuentemente con una periodicidad mensual y, por lo tanto, es importante cuantificar el grado de similitud entre el comportamiento de un mes respecto al anterior. Las secciones electorales

¹¹ Véanse Fernández-Martínez *et al.* (2018), y Heath y Borowski (2013) para una explicación de la definición matemática del coeficiente de Variación Proporcional.

con un nivel de autocorrelación bajo presentarían un mayor nivel de complejidad electoral.

Análisis exploratorio de datos y recursos tecnológicos

Una vez definido el conjunto de indicadores base, se realiza un análisis exploratorio de datos para estos indicadores, con el propósito de caracterizar su comportamiento individual y evaluar la pertinencia de su inclusión para la cuantificación de la complejidad electoral.

Este estudio exploratorio se enfoca en la realización de un análisis de correlación (por medio de la elaboración de correlogramas y mapas de calor) para identificar las interrelaciones en los indicadores base. Asimismo, se efectúa 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.

Respecto a las herramientas computacionales utilizadas para el análisis exploratorio de la información y para la construcción de los ICES y la TCES, destacan las siguientes: el lenguaje de programación Python; la librería Scikit-learn (especializada en algoritmos de aprendizaje automático), las librerías Seaborn y Plotly (para la elaboración de gráficos y animaciones); y, el programa QGIS (para la elaboración de mapas).

La metodología de los Índices de Complejidad Electoral Seccionales

El paradigma actual de la elaboración de indicadores socioeconómicos 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.

Por consiguiente, para obtener un conjunto de métricas que cuantifiquen la complejidad del sistema geoelectoral a escala seccional, y que se apegue al paradigma actual de elaboración de indicadores socioeconómicos, se construye un conjunto de Índices de Complejidad Electoral Seccionales (ICES) sustentados en la información contenida en los siete indicadores base.

Los ICES se pueden definir formalmente como *números índice (es decir, números reales) que valoran el grado de complejidad electoral presente en una unidad territorial, a partir de la síntesis (ponderación) de la información contenida en un grupo diverso de indicadores.*¹² Es decir, los ICES cuantifican en un solo valor numérico el grado de complejidad electoral asociado al comportamiento de los factores geoespaciales, demográficos y estructurales.

Considerando la anterior definición, se deduce que los ICES 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 secciones electorales en función de su grado de complejidad electoral. Asimismo, los ICES también posibilitan el análisis temporal de la complejidad electoral (si, por ejemplo, se recalcularan los ICES en un periodo de tiempo posterior), ya que permiten cuantificar las variaciones en la complejidad electoral ocasionadas por cambios en el comportamiento de los indicadores base.

Para la elaboración de los ICES es necesario utilizar una técnica de reducción de dimensionalidad, ya que, a partir de un 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 ICES se utiliza la técnica de Análisis de Componentes Principales (ACP), ya que 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

¹² Para el caso específico de esta investigación, se hace referencia al grupo de siete indicadores base.

datos para la obtención de índices sintéticos.¹³ Algunas de las características del ACP son las siguientes (véanse Jackson, 1991; Abdi y Williams, 2010):

1. Reducir la dimensionalidad de grandes conjuntos multivariados de datos a través de un proceso de transformación lineal que mapea el conjunto original de variables a un nuevo conjunto reducido de variables llamados Componentes Principales (CP). Es decir, permite tener un nuevo conjunto de datos más conciso que sintetiza las características y patrones esenciales del conjunto original.
2. Estimar en una forma no arbitraria los coeficientes de ponderación de las variables originales que conforman a los CP, y también permite retener y reflejar el máximo de información en un subespacio dimensional reducido.
3. Simplificar el análisis y la interpretación visual de los conjuntos de datos de alta dimensionalidad, ya que se pueden utilizar los primeros CP para proyectar el conjunto de datos a un subespacio de variables más reducido e interpretable.
4. Generar indicadores concisos que permiten tipificar los datos. Específicamente, se puede utilizar el primer CP como índice de ordenación cardinal, ya que esta nueva variable explica la mayor parte de la varianza encontrada en los datos.

Por consiguiente, el ACP es un método que permite manejar la multidimensionalidad del fenómeno estudiado, a la vez que proporciona, mediante la consideración de los CP, un conjunto de índices que sintetizan las características esenciales del conjunto original de indicadores.¹⁴

Para la presente investigación, la valoración que obtenga cada sección electoral en las CP se tomará como el valor correspondiente de los ICES. A partir de los resultados de los ICES —y específicamente, a partir de los

¹³ Véase Shin *et al.* (2020) o Consejo Nacional de Población (Conapo) (2002) para un ejemplo de implementación del Análisis de Componentes Principales, en un problema de indicadores sociales.

¹⁴ Para una exposición más completa sobre los métodos de proyección lineales (como el ACP), véanse Jackson (1991), Abdi y Williams (2010), Hastie y Stuetzle (1989) y Jolliffe (2002).

resultados del ICES asociado al primer Componente Principal (ICES principal)— se puede establecer un ordenamiento de complejidad electoral para el conjunto nacional de secciones electorales.

A pesar de las ventajas que ofrece el ACP para reducir la dimensionalidad y, por ende, para proyectar los datos en un subespacio de indicadores más consistente, es importante examinar los resultados obtenidos con dicho método, para valorar la capacidad efectiva de los CP en sintetizar la información de los indicadores originales.¹⁵

Como se mostrará después, los resultados de la utilización del método ACP indican que no se puede usar únicamente el primer CP para generar un ordenamiento de complejidad electoral consistente, ya que éste no alcanza un nivel adecuado de síntesis de información para los indicadores originales. Es decir, dadas las características de los indicadores que conforman los ICES, no está plenamente justificado utilizar sólo el primer CP como indicador de comparación de la complejidad electoral que presenta cada sección electoral.

Sin embargo, se puede solventar la falta de solidez estadística del primer CP mediante el uso simultáneo de la información contenida en el conjunto total de CP (o en un subconjunto de CP que logre una mayor explicación de la varianza observada en los indicadores originales). Es decir, la proyección de las secciones electorales en el espacio de transformación de uno o más CP permitirá elaborar una Tipología de Complejidad Electoral Seccional (TCES), la cual ofrecerá un método de clasificación más robusto de la complejidad, en comparación a un ordenamiento de complejidad sustentado únicamente en el ICES principal.

La metodología de la Tipología de Complejidad Electoral Seccional

La creación de la TCES implica resolver un problema de agrupamiento no supervisado, ya que el conjunto de datos provenientes de los ICES no se en-

¹⁵ En el ACP, la razón de varianza explicada es la métrica más utilizada para examinar la capacidad de síntesis de las CP.

cuentra etiquetado, es decir, carece de variables objetivo.¹⁶ Por consiguiente, para la creación de la TCES se analizaron diferentes algoritmos, con base en las características —y el volumen— de los datos implícitos en los ICES¹⁷

Favorablemente, la única familia de algoritmos de agrupamiento que no mostró restricción con las capacidades de cómputo disponibles fueron los algoritmos de tipo particional, específicamente el método K-medias. Consecuentemente, para la elaboración de la TCES, se utilizó un método de agrupamiento K-medias para clasificar a las secciones electorales en función de su respectiva valoración en los ICES.

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 pertenece a un solo subgrupo.¹⁸ Para determinar el número óptimo de subgrupos es necesario realizar un análisis de *silhouette score* (también conocido como análisis de coeficientes de silueta), ya que este análisis permite analizar en forma gráfica (a través del llamado diagrama de silueta), la distancia de separación entre los grupos de una configuración específica,¹⁹ la cual se determina por medio de los coeficientes de silueta.²⁰

Por ejemplo, si una sección electoral (perteneciente a un grupo específico) presenta un coeficiente de silueta cercano a +1, esto indica que dicha sección se encuentra muy alejada (en el espacio CP) de otras secciones electorales pertenecientes a grupos vecinos. En el caso de un coeficiente de silueta igual a 0, significa que la sección se encuentra en el límite de decisión entre dos grupos vecinos. Finalmente, para el caso de una sección con un coeficiente de silueta cercano a -1, implica que dicha sección se encuentra

¹⁶ Véase una exposición de los algoritmos de agrupamiento en Dean (2014) y Everitt *et al.* (2011). Asimismo, para una demostración de la implementación de los algoritmos de aprendizaje no supervisado en Python, ver la documentación de la librería Scikit-learn (https://scikit-learn.org/stable/auto_examples/index.html).

¹⁷ 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 probó sin éxito con diferentes algoritmos de agrupamiento basados en modelos (Gaussian Mixture, DBSCAN y OPTICS).

¹⁸ 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.

¹⁹ Existen diferentes tipos de métricas de distancia. En esta investigación se decidió utilizar la métrica euclidiana para establecer las distancias entre los puntos de datos, ya que su interpretación es más intuitiva.

²⁰ Estos coeficientes tienen un rango entre [-1, +1].

más cercana a alguna sección perteneciente a otro grupo, en comparación con las secciones que se encuentran dentro de su mismo grupo de pertenencia, es decir, dicha sección está asignada a un grupo incorrecto.²¹

Resultados de los Índices de Complejidad Electoral Seccionales

La construcción de los ICES inicia con el análisis de correlación de los siete indicadores base. Posteriormente, se implementa la técnica ACP para reducir la dimensionalidad del conjunto de indicadores base, y obtener las correspondientes CP. Una vez calculados los CP, se analiza la capacidad de síntesis de información para cada uno de ellos, y se determina el número óptimo de CP que conformará el grupo de ICES. Finalmente, se obtiene la proyección tridimensional de las secciones electorales en el nuevo espacio de transformación de los ICES y se analiza la emergencia de patrones distinguibles en las proyecciones.

Análisis exploratorio y construcción de los ICES

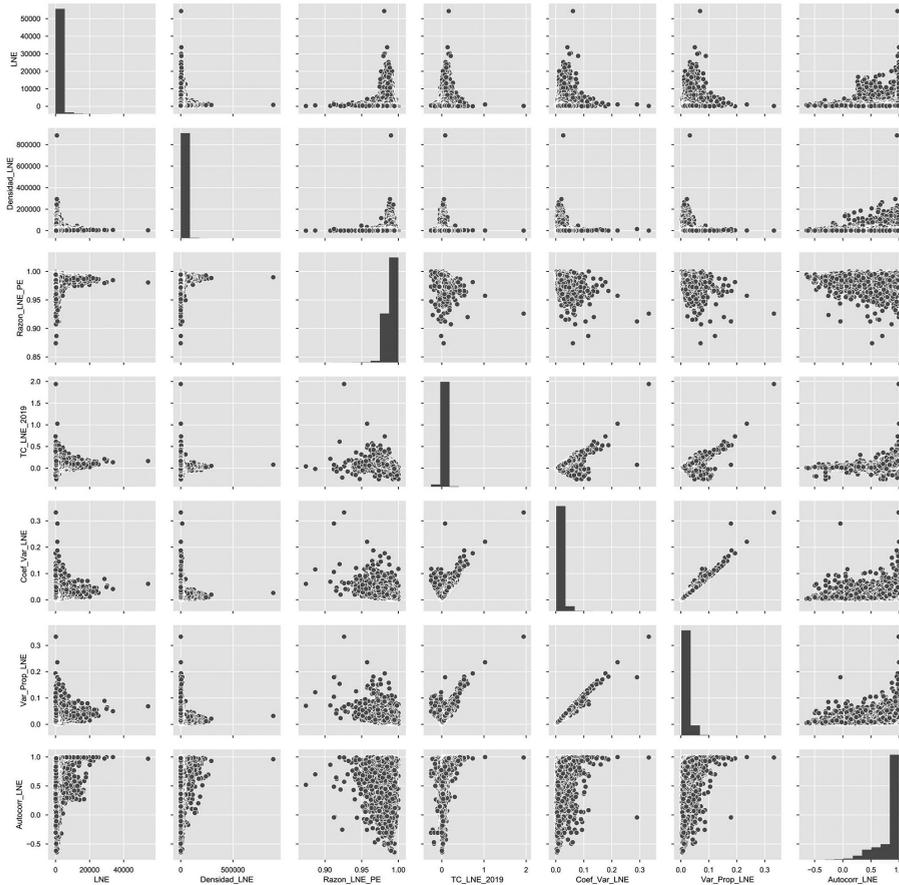
En la figura 1 se presentan los resultados del análisis de correlación de los siete indicadores base y la distribución que presenta cada indicador (histogramas sobre la diagonal del panel).²² Se puede observar que los siete indicadores (a excepción del indicador Razon_LNE_PE) presentan distribuciones asimétricas, lo cual denota que hay un número significativo de secciones que poseen características geoespaciales, demográficas y estructurales atípicas respecto a la mayoría de las secciones.²³

²¹ Para una explicación técnica y detallada del funcionamiento del análisis de silueta se recomienda ver Rousseeuw (1987).

²² Cada punto en los diagramas de dispersión representa alguna de las 61 004 secciones electorales analizadas.

²³ Para una mejor comprensión de este panel, considérese la gráfica localizada en la primera fila y columna (histograma del indicador LNE). Este histograma muestra que la mayoría de las secciones electorales tiene un número reducido de ciudadanos inscritos en la LNE, y esto se debe primordialmente a la normatividad de mantener un límite máximo de 3000 ciudadanos por sección. Sin embargo, el histograma también muestra que hay algunas secciones que rebasan el límite establecido.

Figura 1. Panel de gráficas de correlación y distribución de los indicadores que conforman los ICES

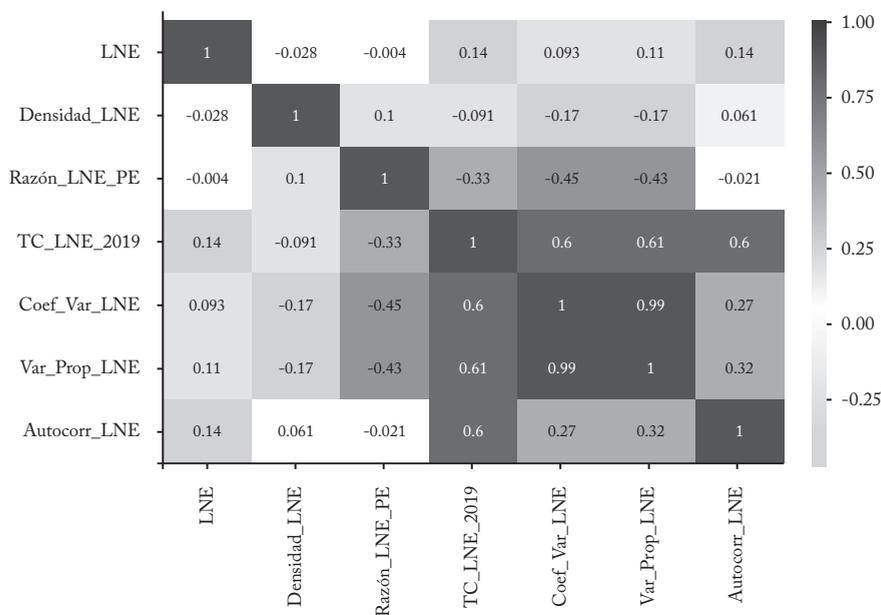


Fuente: Elaboración propia con la librería Python-Pandas-Seaborn.

Los resultados muestran la presencia de correlaciones significativas entre los indicadores, que se visualizan mejor en el mapa de calor de la figura 2. En conjunto, ambas gráficas señalan que los indicadores asociados a la variabilidad demográfica (TC_LNE_2019, Coef_Var_LNE y Var_Prop_LNE) tienen correlaciones positivas entre sí. En contraste, el indicador asociado

al nivel de actualización de la LNE y el PE (Razon_LNE_PE) se encuentra correlacionado negativamente con los indicadores asociados a la variabilidad demográfica. Consecuentemente, los resultados del análisis de correlación justifican la aplicación del ACP al conjunto de indicadores base.²⁴

Figura 2. Mapa de calor de la matriz de correlación de los siete indicadores seccionales base

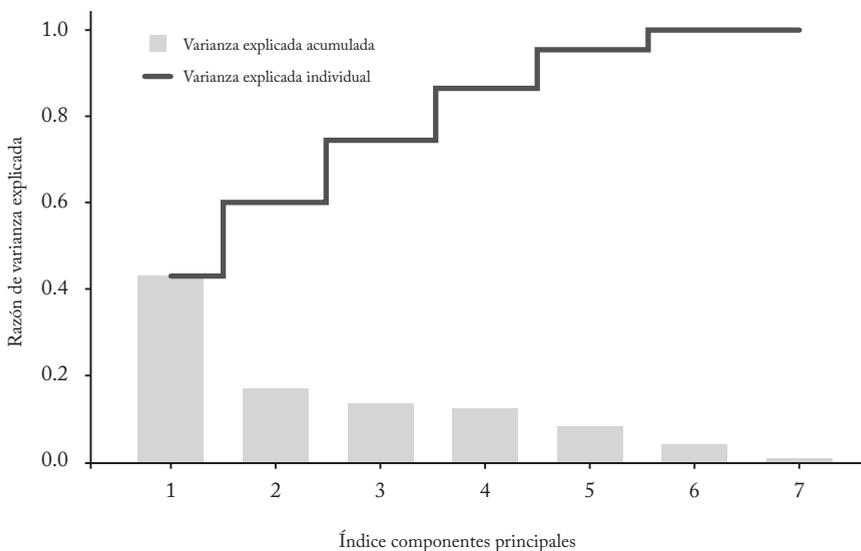


Fuente: Elaboración propia con la librería Python-Pandas-Seaborn.

²⁴ Se recomienda revisar Abdi y Williams (2010) para una explicación detallada de las condiciones que deben cumplir los datos para obtener una adecuada transformación con el ACP.

La primera fase de la implementación del ACP es la estandarización del conjunto de indicadores base (normalizar y centralizar).²⁵ Una vez realizada esta estandarización, se calcula la matriz de covarianza, junto con sus respectivos *eigenvalores* y *eigenvectores*, siendo estos últimos los CP seccionales. El resultado de la varianza explicada por cada CP seccional se muestra en la figura 3.

Figura 3. ICES: proporción de varianza explicada por Componente Principal



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

Se puede observar que el primer Componente Principal (CP1) alcanza un porcentaje de varianza explicada de 43 %; consecuentemente, el CP1 se considera el principal ICES. Sin embargo, el CP1 no es un indicador que

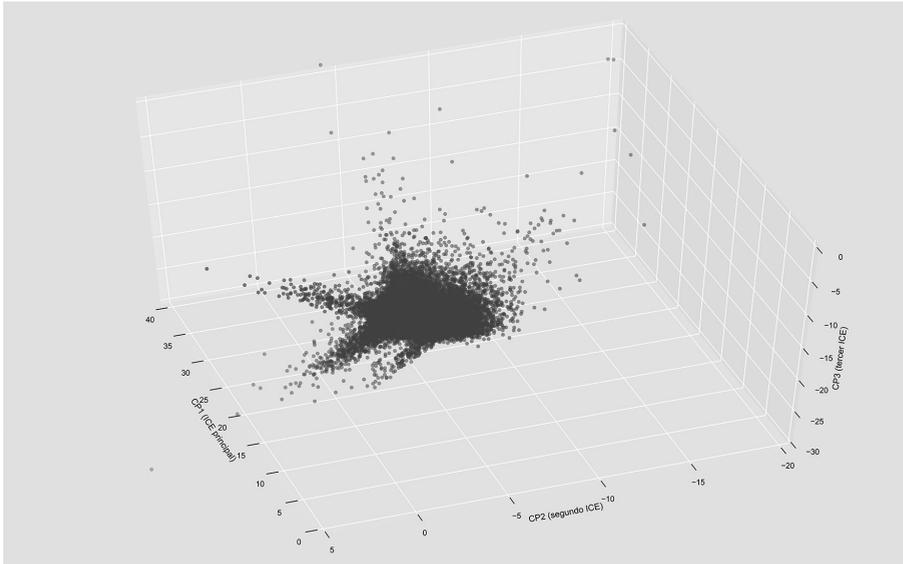
²⁵ El propósito de la estandarización es reducir la varianza observada en los indicadores base, ya que éstos se encuentran en diferentes unidades de medida, lo cual induce una falsa variabilidad en los datos.

syntheticamente las características de los siete indicadores base. Para realizar comparaciones consistentes de complejidad entre las secciones electorales, el CP1 debe utilizarse en comunión con los subsiguientes Componentes Principales.

Es importante señalar que el segundo y tercer Componente Principal (CP2 y CP3) tienen un nivel de explicación de la variabilidad de los datos de 17 y 14 %, respectivamente. En consecuencia, el conjunto de tres CP conforman el grupo de ICES.²⁶ Estos resultados justifican la realización de una tipología de secciones electorales con base en las componentes CP1, CP2 y CP3, ya que, en forma individual, cada CP (es decir, cada ICES) no alcanza un nivel satisfactorio de síntesis de la información.

²⁶ La elección de cuántos Componentes Principales deben elegirse para proyectar las unidades territoriales en un espacio dimensional reducido es hasta cierto punto subjetiva, ya que depende del nivel de síntesis de información que se busque alcanzar. Por lo tanto, considerando los objetivos y las características de los indicadores analizados, se decide elegir los primeros tres Componentes Principales como los respectivos ICES.

Figura 4. Proyección de las secciones electorales en el espacio CP1-CP2-CP3



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

En la figura 4 se proyecta el espacio completo de las componentes CP1-CP2-CP3; y se identifican patrones de agrupamiento en la dispersión de las secciones electorales (véanse los picos de la nube de dispersión).²⁷ A partir de estos patrones se infiere la existencia de grupos de secciones electorales con características similares, o anómalas respecto al comportamiento promedio observado en el conjunto total de secciones electorales.

²⁷ Para visualizar con mayor detalle los patrones de agrupamiento de las secciones electorales en el espacio de los CP, ver la animación [Pyplotly_Proyeccion3d_Seccc.html](#).

Interpretación de los ICES

Se observan en la tabla 1 los indicadores que mayor peso tienen en la CP1 (ICE principal): el Coeficiente de Variabilidad Proporcional, el Coeficiente de Variación y la Tasa de Crecimiento de la LNE. En el segundo Componente Principal CP2 (segundo ICE), los indicadores Autocorrelación de la LNE y Densidad de LNE tienen más peso. Finalmente, en la tercera componente CP3 (tercer ICE) tienen mayor importancia los indicadores de Densidad de LNE y Magnitud de LNE.

Dado que los indicadores que más pesan en el ICE principal están relacionados con el crecimiento y las variaciones poblacionales observadas en la LNE, este índice describe la complejidad electoral asociada a la variabilidad demográfica. Es decir, entre mayor sea la valoración de una sección electoral en el ICE principal, mayor será su variabilidad demográfica.

Tabla 1. ICES: composición de los CP en función de los indicadores base

Indicador	CP1 (ICE principal)	CP2 (segundo ICE)	CP3 (tercer ICE)
Magnitud de LNE	0.11031145	0.35273595	-0.7648855
Densidad de LNE	-0.1179742	0.46068761	0.60983702
Razón LNE entre Padrón Electoral	-0.31617025	0.40298392	-0.1649244
Tasa de Crecimiento LNE (2019)	0.47102119	0.26156096	0.05813301
Coeficiente de Variación de la LNE	0.52698938	-0.17726434	0.05273008
Coeficiente de Variabilidad Proporcional de la LNE	0.53223215	-0.13808495	0.04144545
Autocorrelación de la LNE (1 retraso)	0.3018278	0.61809708	0.08928204

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

Por otra parte, el segundo ICE está influenciado primordialmente por el nivel de Autocorrelación de la LNE respecto al comportamiento mensual previo. Este índice se puede interpretar como una métrica de la correlación entre el comportamiento reciente de la LNE con el comportamiento futuro de corto plazo. Asimismo, los indicadores Densidad de LNE y Razón LNE/PE también influyen positivamente en el segundo ICE. En consecuencia, entre mayor sea la valoración de una sección electoral en el segundo ICE, mayor es la facilidad de pronosticar el comportamiento de corto plazo de su LNE, el nivel de actualización de la LNE/PE y la densidad poblacional.

Finalmente, el tercer ICE presenta dos influencias principales. La primera influencia (ponderador positivo) viene del indicador de Densidad LNE y la segunda (ponderador negativo) proviene del indicador de Magnitud LNE. Existe una correlación directa entre ambos indicadores, ya que la densidad es igual al indicador de magnitud de LNE multiplicado por un valor escalar (superficie de la sección). Consiguientemente, las secciones electorales que obtengan una valoración alta en el tercer ICE se caracterizan por tener un bajo número de ciudadanos inscritos en la LNE, o por tener una reducida superficie urbanizada (y, por ende, una alta densidad).²⁸

Sería importante analizar la valoración que obtuvo cada sección electoral en los tres ICES. Sin embargo, la cantidad de secciones electorales analizadas (61 004 secciones) impide ofrecer en este documento un desglose completo de los resultados obtenidos con los tres ICES.²⁹

Resultados de la Tipología de la Complejidad Electoral Seccional

Esta sección presenta los resultados de la implementación del algoritmo de partición K-medias para construir la TCES. El algoritmo utiliza los re-

²⁸ Para complementar las explicaciones y resultados de esta sección, se incluye en el repositorio de GitHub el código de Python utilizado para la implementación del ACP. Asimismo, se encuentra una animación de la proyección de las secciones en el espacio de los CP (Pyplotly_Proyeccion3d_Secc.html), y un estadístico con el listado completo de las secciones electorales con su respectiva valoración en los ICES (Resultados_Secc_Indic_pca_Kmeans_ind7.csv).

²⁹ Para observar el desglose completo del *ranking* de complejidad electoral, véase el estadístico Resultados_Secc_Indic_pca_Kmeans_ind7.csv, incorporado en el repositorio de GitHub.

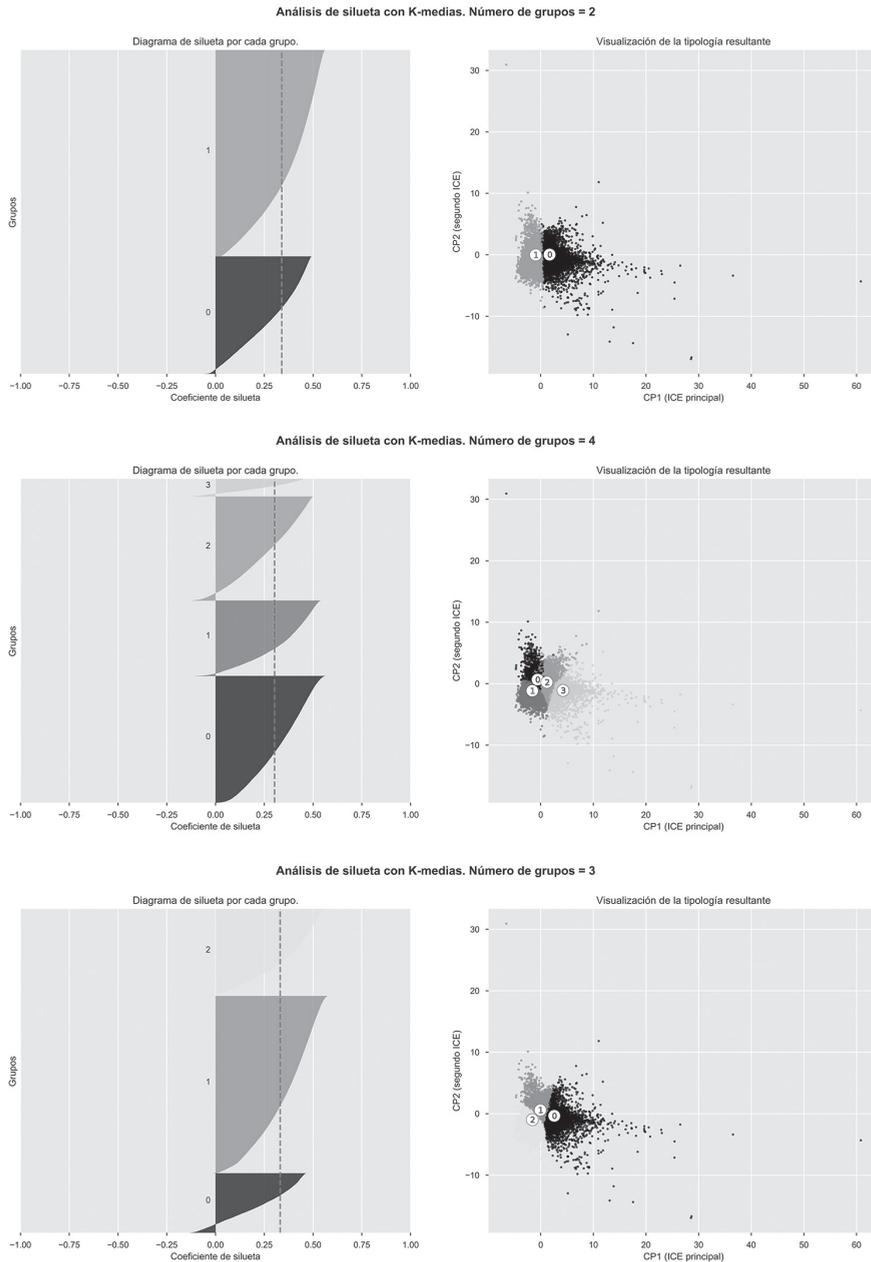
sultados de la proyección tridimensional de las secciones electorales en el espacio de los ICES para realizar un análisis de agrupamiento, el cual permite formar grupos de secciones con grados de complejidad electoral similar. Asimismo, se realiza un análisis de silueta para determinar la configuración óptima de grupos en la TCES. Enseguida, se clasifica el conjunto total de secciones electorales en función de la TCES y se analizan las estadísticas de tendencia central de cada grupo de la tipología para identificar los rasgos característicos de cada grupo —en términos de los siete indicadores base—, así como los casos atípicos de complejidad electoral.

Análisis de silueta e implementación de K-medias

En la figura 5 se muestran los diagramas de silueta correspondientes a las tres posibles configuraciones de grupos analizadas (configuraciones de 2, 3 y 4 grupos).³⁰ Los resultados observados en los diagramas de silueta señalan que las configuraciones de 2 y 4 grupos no son óptimas, ya que muestran un número significativo de secciones electorales agrupadas incorrectamente, además de que el tamaño de los grupos es significativamente inequitativo. En contraste, la configuración de tres grupos es óptima, ya que ofrece una distribución más equitativa de las secciones electorales en los grupos de la tipología. Por lo tanto, considerando los resultados obtenidos a partir del análisis de coeficientes de silueta, se selecciona la configuración de tres grupos para conformar la TCES.

³⁰ Para facilitar la visualización de los diagramas de silueta, y el correspondiente agrupamiento en el espacio de los CP, sólo se grafica la representación bidimensional.

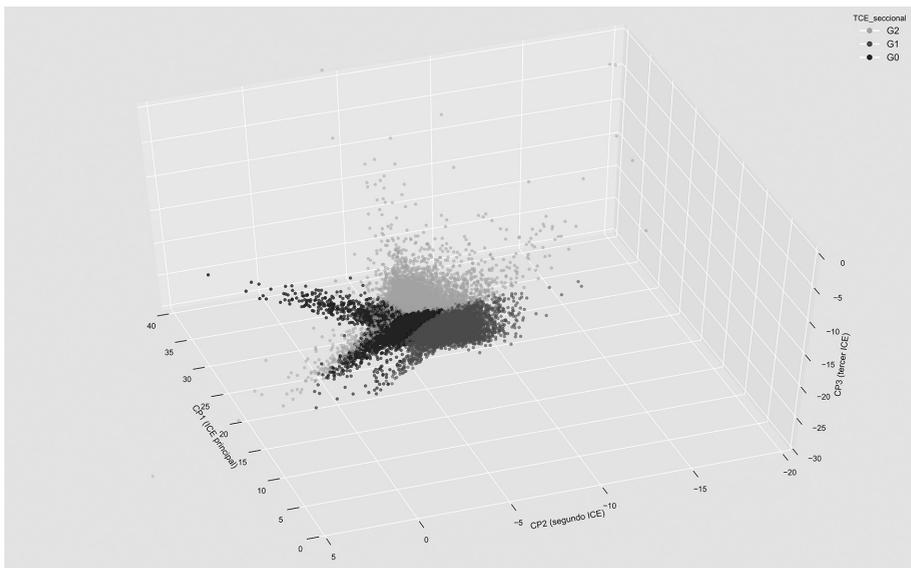
Figura 5. Diagramas de silueta para las configuraciones de 2, 3 y 4 grupos



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

En la figura 6 se presenta el resultado de la TCES obtenida con el algoritmo K-medias (en el espacio CP1-CP2-CP3), utilizando la configuración óptima de tres grupos (etiquetas G0, G1, G2). En esta figura se puede observar el ajuste de la tipología seccional a la estructura tridimensional de la proyección, ya que la TCES 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 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 TCES separa las secciones electorales que tienen una valoración positiva en el segundo ICE (grupo G0 y, en menor medida, el grupo G2) de las secciones que presentan una valoración negativa (secciones pertenecientes al grupo G1).³¹

Figura 6. Proyección de las secciones electorales, con agrupamiento por K-medias, en el espacio CP1-CP2-CP3



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

³¹ Se recomienda ver el archivo `Pyplotly_Proyeccion3d_Secc.html` para visualizar, en forma interactiva, la proyección de las secciones electorales en el espacio de los ICES, junto con sus correspondientes clasificaciones en la TCES.

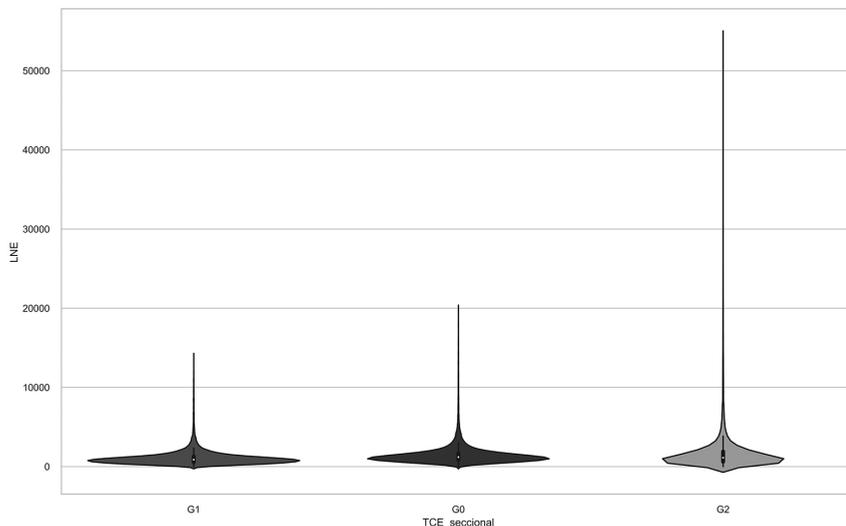
Por último, en el anexo 2 (mapas) se tiene la distribución geoespacial de las secciones electorales con su respectiva clasificación en la TCES.

Estadísticas de los grupos de la TCES

Aquí se analizan y comparan —mediante gráficas de violín— las distribuciones y estadísticas de tendencia central de los siete indicadores base y de los tres ICES. Para la estimación de las estadísticas de tendencia central se agruparon las secciones electorales en función de su clasificación en la TCES. 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 atípico.

Con el propósito de ayudar a la comprensión del panel de gráficas de violín (figura 8), se analiza en la figura 7 el caso particular de la gráfica de violín del indicador LNE.

Figura 7. Gráfica de violín del indicador LNE (con agrupación en función de la TCES)



Fuente: Elaboración propia a partir de la librería Python-Scaborn.

En la figura 7 se observa que los grupos G0, G1 y G2 exhiben estadísticas de tendencia central similares; los tres grupos muestran medias de LNE equiparables (indicadas por los puntos blancos dentro de las figuras de violín), que también poseen rangos intercuartílicos (representados por los rectángulos negros) e intervalos de confianza (líneas verticales) similares.

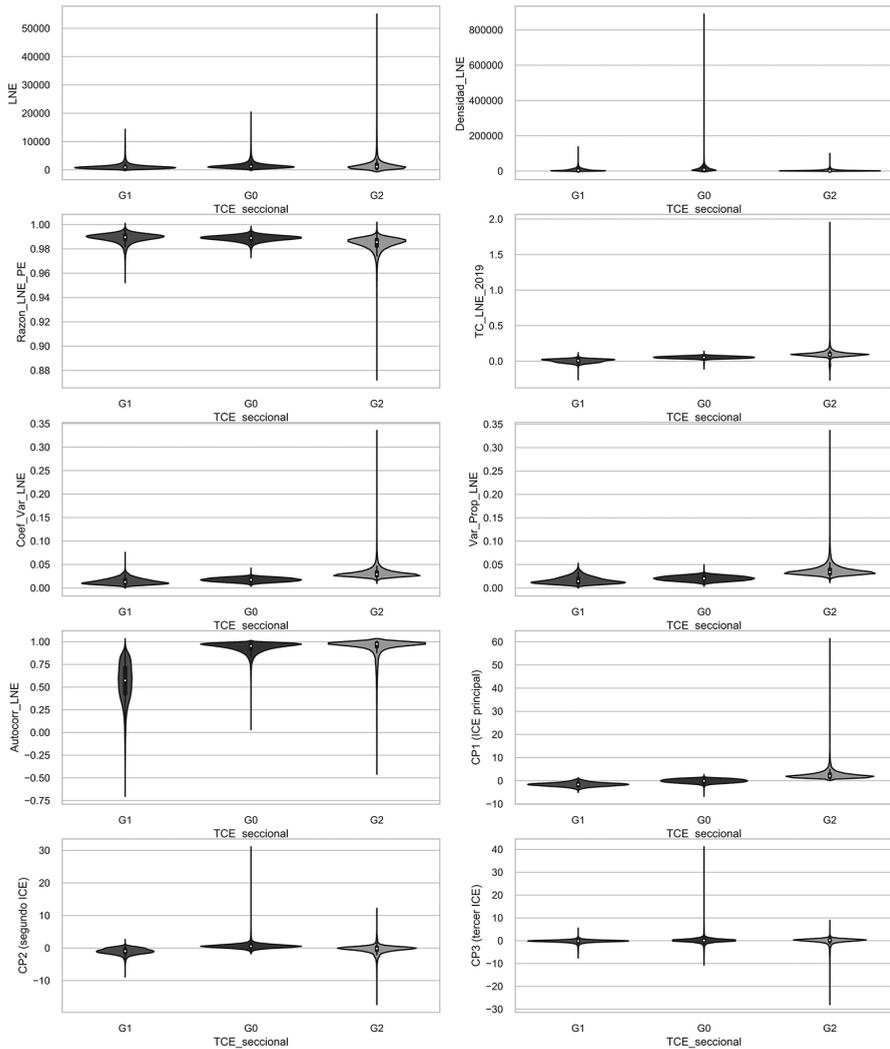
Por otra parte, la figura 7 también muestra las distribuciones de valores (LNE) correspondientes a cada grupo de la TCES. En particular, se puede observar que los tres grupos de la TCES presentan valores atípicos (figuras de violín alargadas), especialmente el grupo G2, ya que éste exhibe una distribución de comportamiento leptocúrtico.³²

³² Una distribución más pronunciada alrededor de la media y con colas más gruesas que la distribución normal.

Considerando la interpretación de la información contenida en la figura 7, se procede al análisis del panel de diagramas de violín mostrado en la figura 8. El panel muestra que los tres grupos de la TCES se caracterizan por presentar estadísticas similares en cada uno de los siete indicadores base. La única excepción es el indicador `Autocorr_LNE`, ya que en éste se pueden apreciar diferencias significativas entre las estadísticas de tendencia central del grupo G1, y las estadísticas de tendencia central de los grupos G0 y G2.

Por otra parte, las distribuciones que se observan en cada indicador muestran ligeras diferencias comparativas en los tres grupos de la tipología seccional (el único caso excepcional es el indicador `Autocorr_LNE`, en el cual el grupo G1 muestra una distribución bastante disímil respecto a las distribuciones observadas en los grupos G0 y G2). En particular, el panel indica que todas las distribuciones de los indicadores base son asimétricas, con diferentes grados y direcciones de sesgo.

Figura 8. Panel de gráficas de violín de los siete indicadores base, más los ICES (con agrupación en función de la TCES)



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

Respecto a la presencia o ausencia de colas largas en las distribuciones de los indicadores (agrupados en función de la TCES), la figura 8 permite apreciar distintos casos. Por ejemplo, en el caso del grupo G2, las distribuciones de los indicadores muestran un comportamiento leptocúrtico, ya que hay una alta cantidad de secciones electorales fuera de rango (valores atípicos), especialmente en los indicadores relacionados con la variabilidad demográfica. En el caso del grupo G0, se observa una distribución con un marcado comportamiento leptocúrtico en los indicadores de Autocorr_LNE y Densidad_LNE (debido a la cantidad significativa de secciones electorales fuera de rango), situación que se repite en el grupo G1, ya que también se observa un marcado comportamiento leptocúrtico en el indicador Autocorr_LNE.

Asimismo, es importante analizar las estadísticas y las distribuciones observadas en los ICES.³³ De acuerdo con la figura 8, las distribuciones de los tres ICES muestran un número significativo de secciones electorales fuera de rango (valores atípicos) que rompen la simetría de las distribuciones. El origen de estos valores atípicos se puede rastrear en las distribuciones que mostraron los grupos G0 y G2 en los indicadores relacionados con la variabilidad demográfica y también con el indicador de Densidad_LNE. Específicamente, las distribuciones del segundo y tercer ICES (en el grupo G0) se pueden explicar por la distribución leptocúrtica observada en los indicadores Densidad_LNE y Autocorr_LNE, mientras que las distribuciones del principal, segundo y tercer ICES (en el grupo G2), se pueden explicar por las distribuciones leptocúrticas presentes en los indicadores base relacionados con la variabilidad demográfica.

Finalmente, es importante destacar la situación atípica de algunas secciones electorales, ocasionada por la valoración que obtuvieron en los ICES, o por la falta de información en alguno de los siete indicadores base. Dichas secciones electorales son relevantes de monitorear, ya que presentan características demográficas y geoespaciales que aumentan significativamente su nivel de complejidad electoral, o presentan inconsistencias en su informa-

³³ Diagramas de violín ubicados en la penúltima y última fila de la figura 8.

ción. En la siguiente subsección se abunda más sobre las características de estas secciones electorales atípicas, lo cual permite elaborar algunas recomendaciones.³⁴

*Identificación de casos atípicos*³⁵

Los resultados de la TCES señalan tres grupos de secciones electorales atípicas.³⁶ El primer grupo está conformado por secciones clasificadas dentro del grupo G2 y que presentan valores atípicos en los tres ICES. Estas secciones se caracterizan por una alta variabilidad demográfica (altas tasas de crecimiento de la LNE), lo cual tiene una incidencia directa en el nivel de actualización de la LNE y el PE (valoración baja en el indicador Razón_LNE/PE) y en la dificultad de pronosticar la evolución mensual de la LNE (baja autocorrelación y altos coeficientes de Variación y Variabilidad Proporcional). En consecuencia, para este grupo se sugiere monitorear la variabilidad demográfica de la LNE, ya que es probable que, en el corto plazo, las secciones electorales de este grupo rebasen el límite demográfico de 3000 ciudadanos por sección. Por otro lado, para la elaboración de pronósticos demográficos, se deben considerar métodos y modelos idóneos para manejar una alta variabilidad demográfica en la LNE.

El segundo grupo de secciones electorales atípicas está compuesto por secciones clasificadas en el grupo G0, con valores atípicos en los tres ICES. Estas secciones se caracterizan por tener altos niveles en la magnitud y densidad de la LNE, así como por presentar niveles bajos de autocorrelación. Esto permite asumir que estas secciones se encuentran ubicadas en

³⁴ En el anexo 2 se presenta un mapa con la ubicación geoespacial de las secciones electorales atípicas.

³⁵ Para la identificación de las secciones atípicas, se utilizó la regla del rango intercuartílico. Este método es directo y sencillo de aplicar usando los cuartiles de las distribuciones, por grupo, de los ICES. Sin embargo, es importante puntualizar que la presencia misma de los valores atípicos puede alterar significativamente su identificación. En esta investigación no se utilizan o comparan métodos más sofisticados para la identificación de valores atípicos (ésta puede ser una línea de investigación futura), pero se recomienda revisar la exposición de Jackson y Chen (2004) para obtener un panorama general de dichos métodos.

³⁶ En los archivos anexos del repositorio de GitHub (Resultados_Secce_Indic_pca_Kmeans_ind7.csv y Pyplotly_Proyeccion3d_Secce.html) se hace un desglose de las secciones electorales con su respectiva valoración en los ICES, su agrupación en la TCES y su tipo de atipicidad.

zonas urbanas y fuera del límite demográfico de 3000 ciudadanos. Otra característica de estas secciones es su estabilidad desde el punto de vista demográfico, ya que hay poca variabilidad en su LNE. Por ende, para este grupo se propone considerar un programa de reseccionamiento, ya que se encuentran fuera del límite demográfico de 3000 ciudadanos.

El tercer grupo de secciones atípicas está conformado por secciones clasificadas dentro del grupo G1, las cuales presentan valores atípicos en los tres ICES. Este grupo se caracteriza por tener una alta LNE, altos niveles de desactualización de la LNE/PE y niveles negativos de autocorrelación. Los resultados indican que estas secciones se encuentran en zonas suburbanas —por la baja densidad de su LNE, lo cual implica una superficie territorial amplia— fuera del límite demográfico de 3000 ciudadanos. Sin embargo, estas secciones son estables desde el punto de vista demográfico, ya que hay poca variabilidad en su LNE.

En relación con las secciones que no tienen una valoración en los ICES —y, por ende, tampoco poseen una clasificación en la TCES— se identifica un patrón en su distribución geoespacial, ya que la mayoría de ellas se encuentra localizada en las zonas rurales de los estados de Chihuahua, Coahuila, Durango, Hidalgo, Michoacán, Nayarit y Sinaloa.³⁷

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_LNE. Para estimar dicho indicador es necesario contar con los datos de amezanamiento 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 tiene registrado su amezanamiento o porque son zonas rurales que efectivamente no cuentan con un amezanamiento discernible.

³⁷ En el anexo 2, véase la figura 2.

Conclusiones

La presente investigación contribuye al estudio de los sistemas sociales complejos, en particular al de los sistemas geoelectorales. El trabajo ha construido instrumentos analíticos (los ICES y la TCES) para cuantificar la complejidad del sistema geoelectoral mexicano a escala seccional; instrumentos que le permitirían al INE monitorear y evaluar eficazmente el RFE.³⁸

Asimismo, se elaboró un marco metodológico original fundamentado en tres elementos: el concepto de complejidad electoral, el conjunto de indicadores base, y en las técnicas de análisis multivariante y minería de datos para la construcción de los ICES y la TCES.

El concepto de *complejidad electoral* se definió como el conjunto de factores geoespaciales, demográficos y estructurales que dificultan u obstaculizan las labores de integración, mantenimiento y actualización del RFE. Los indicadores base otorgaron una descripción más precisa de las actuales características geoespaciales, demográficas y estructurales presentes en el sistema geoelectoral mexicano. El ACP proporcionó un método para manejar la multidimensionalidad implícita en el conjunto de indicadores base; tal que se modeló, con un menor número de variables (los CP), la información contenida en estos indicadores.

Por otra parte, los diagramas de varianza explicada mostraron las capacidades de síntesis de información de cada uno de los CP. Los resultados permiten concluir que los CP, considerados individualmente, presentan algunas limitaciones para sintetizar adecuadamente la información contenida en los indicadores base; sin embargo, cuando se consideran en conjunto se tiene un nivel de síntesis de información superior al obtenido en forma individual por cada CP. En consecuencia, el conjunto de ICES se forma a partir de los primeros tres CP.

³⁸ Asimismo, estos instrumentos también proporcionan información adicional sobre la dinámica demográfica local, lo cual puede sentar las bases para la elaboración de un sistema de monitoreo demográfico a nivel nacional, ya que los ICES pueden actualizarse continuamente tan pronto se genere nueva información proveniente de las tareas de integración de la LNE y del PE.

El análisis de la estructura tridimensional de la proyección de las secciones electorales en el espacio de los ICES arrojó un resultado importante: la identificación de patrones de agrupamiento. Se comprobó la existencia de grupos particulares de secciones electorales con características geoespaciales, demográficas y estructurales similares y, por ende, con similitudes en su grado de complejidad electoral. Así, las estructuras observadas en las proyecciones justificaron la elaboración de una TCES.

El análisis de coeficientes de silueta permitió determinar el número óptimo de grupos a considerar en la tipología, evaluar la pertinencia de usar el algoritmo de agrupamiento K-medias y concluir que la configuración óptima de la TCES es de tres grupos. Y con base en esta configuración de grupos se implementó el algoritmo K-medias en las proyecciones tridimensionales de las secciones en el espacio de los ICES. El resultado final es la clasificación de todas las unidades territoriales básicas que conforman el sistema geoelectoral mexicano.

Con relación a la identificación de los casos atípicos de complejidad electoral, ésta es una aportación significativa, ya que puede influir positivamente en el mejoramiento de la calidad del RFE. El análisis de los casos atípicos permite encontrar grupos específicos de secciones electorales que presentan características anómalas respecto a la evolución de su LNE, o respecto a sus características geoespaciales.

En relación con las limitaciones de los ICES y la TCES se proponen algunas líneas de investigación para superarlas y para expandir los resultados alcanzados.

La metodología de cálculo de los ICES y la TCES se fundamentó primordialmente en el uso del ACP, y en la implementación del método de agrupamiento K-medias. Esta elección 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 de proyección y agrupamiento de datos.

Por consiguiente, se propone para una investigación futura utilizar otras tecnologías de análisis y procesamiento de datos que permitan remontar las limitantes que se encontraron. Por ejemplo, considerando el volumen significativo de secciones, se podrían ampliar las capacidades de procesa-

miento de datos de los métodos de agrupamiento utilizando tecnologías de cómputo distribuido. Asimismo, se podría mejorar la eficiencia de los ICES y la TCES 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 mantenimiento y actualización del RFE.

Otra línea de investigación utilizaría los ICES y la TCES 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 (estructura demográfica, flujos de migración, niveles de marginación, accesibilidad geográfica, etcétera). Sin embargo, la incorporación de indicadores adicionales se encontrará sujeta a la disponibilidad de información. Por consiguiente, se recomienda recabar más información en las escalas básicas del sistema geoelectoral mexicano. Asimismo, se sugerirían actualizaciones periódicas de los ICES y la TCES 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 de los instrumentos registrales.



Fuentes de consulta

- Abdi, Hervé y Williams, Lynne J. (2010). “Principal Component Analysis”. *WIREs Comp Stat*, 2, 433-459. doi:10.1002/wics.101
- 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:10.29101/crcs.v27i83.12148
- Consejo Nacional de Población (2002). *Índices de Marginación a Nivel Localidad*. México: Autor.
- Constitución Política de los Estados Unidos Mexicanos. Recuperado de http://www.diputados.gob.mx/LeyesBiblio/pdf_mov/Constitucion_Politica.pdf
- Dean, Jared (2014). *Big data, data mining, and machine learning: value creation for business leaders and practitioners*. New Jersey: Wiley.
- 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. doi: 10.1002/ecs2.2527
- Flores Valdés, Jorge y Martínez Mekler, Gustavo (2011). *Encuentros con la complejidad*. México: Siglo Veintiuno Editores.
- GitHub. Disponible en https://github.com/MDAlvarezH/Articulo_Codigo-Analisis-sistema-geo-electoral-mexicano-
- Hastie, Trevor y Stuetzle, Werner (1989). “Principal Curves”. *Journal of the American Statistical Association*, 84 (406), 502-516. doi: 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. doi: 10.1371/journal.pone.0084074

- Horgan, John (2015). *The End of Science: Facing the Limits of Knowledge in the Twilight of the Scientific Age*. New York: Addison-Wesley.
- Jackson, Donald A. y Chen, Yong (2004). “Robust Principal Component Analysis and Outlier Detection with Ecological Data. *Environmetrics*, 15, 129-139. doi: 10.1002/env.628
- Jackson, J. E. (1991). *A User’s Guide to Principal Components*. New Jersey: Wiley.
- Jolliffe, I. T. (2002). *Principal Component Analysis*. New York: Springer.
- Ladyman, James; Lambert, James; y Wiesner, Karoline (2013). “What is a Complex System?”. *Euro Jnl Phil Sci*, 3, 33-67. DOI: 10.1007/s13194-012-0056-8
- Ley General de Instituciones y Procedimientos Electorales. Recuperado de <http://www.diputados.gob.mx/LeyesBiblio/ref/lgipe.htm>
- Lloyd, Seth (2001). “Measures of Complexity: A Nonexhaustive List”. *IEEE Control Systems Magazine*, 21 (4), 7-8. doi: 10.1109/mcs.2001.939938
- Maggino, Filomena (2017). *Complexity in Society: From Indicators Construction to their Synthesis*. Switzerland: Springer.
- Mitchell, Melanie (2009). *Complexity: A Guided Tour*. London: Oxford University Press.
- Phelan, Steven (2001). “What is Complexity Science, Really?”. *Emergence*, 3 (1), 120-136. doi: 10.1207/S15327000EM0301_08
- 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. doi: 10.1016/0377-0427(87)90125-7
- Scikit-learn. Disponible en https://scikit-learn.org/stable/auto_examples/index.html
- Shin, J.; Price, M. H.; Wolpert, D. H.; Shima, H.; Tracey, B.; y Kohler, T. A. (2020). “Scale and Information-Processing Thresholds in Holocene Social Evolution”. *Nature Communications*, 11 (2394), 1-8. doi: 10.1038/s41467-020-16035-9
- Waldrop, Mitchel (1992). *Complexity. The Emerging Science at the Edge of Order and Chaos*. New York, EE. UU.: Simon & Schuster.
- Wolfram, Stephen (2002). *A New Kind of Science*. Champaign, EE. UU.: Wolfram Media.

Anexo 1: Fórmulas de los indicadores base

L_1^l ≡ número de ciudadanos inscritos en la LNE de la sección electoral l , al corte de enero de 2019.

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

L_i^l ≡ número de ciudadanos inscritos en la LNE de la sección electoral l , al corte del mes i de 2019.

A_{12}^l ≡ área superficial (kilómetros cuadrados) del amanzanamiento urbano de la sección electoral, al corte de diciembre de 2019.

P_{12}^l ≡ número de ciudadanos registrados en el Padrón Electoral de la sección electoral l , al corte de diciembre de 2019.

N ≡ 12 meses

1. Magnitud de la LNE

$$LNE^l \equiv L_{12}^l$$

2. Densidad de la LNE

$$Densidad_LNE^l \equiv \frac{L_{12}^l}{A_{12}^l}$$

3. Razón LNE entre Padrón Electoral

$$Razon_LNE_PE^l \equiv \frac{L_{12}^l}{P_{12}^l}$$

4. Tasa de Crecimiento Anual de la LNE

$$TC_LNE_2019^l \equiv \frac{L_{12}^l - L_1^l}{L_1^l}$$

5. Coeficiente de Variación de la LNE

$$Coef_Var_LNE^l \equiv \frac{\sigma_L^l}{\bar{L}_{2019}^l} = \frac{\sqrt{\frac{1}{N} \sum_{i=1}^N (L_i^l - \bar{L}_{2019}^l)^2}}{\sum_{i=1}^N \frac{L_i^l}{N}}$$

6. Coeficiente de Variabilidad Proporcional de la LNE

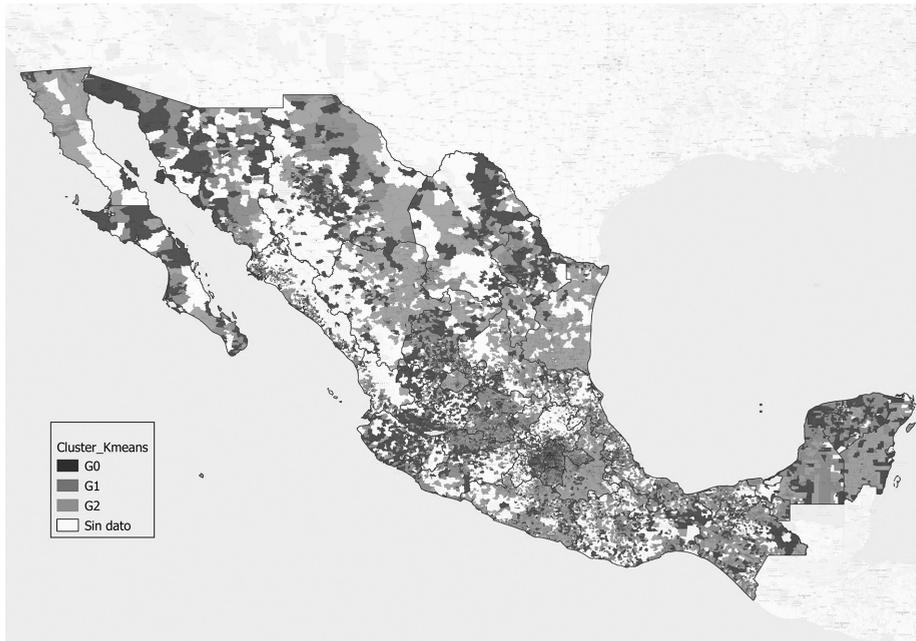
$$Var_Prop_LNE^l \equiv \frac{1}{C} \sum_{comb} D(L_i^l, L_j^l) = \frac{2}{N(N-1)} \sum_{comb} 1 - \frac{\min(L_i^l, L_j^l)}{\max(L_i^l, L_j^l)}$$

7. Autocorrelación de la LNE (k = retraso mensual)

$$Autocorr_LNE^l \equiv r_k^l = \frac{\sigma_k^l}{\sigma_L^l} = \frac{\frac{1}{N} \sum_{i=1}^{N-k} (L_i^l - \bar{L}_{2019}^l) (L_{i+k}^l - \bar{L}_{2019}^l)}{\frac{1}{N} \sum_{i=1}^N (L_i^l - \bar{L}_{2019}^l)^2}$$

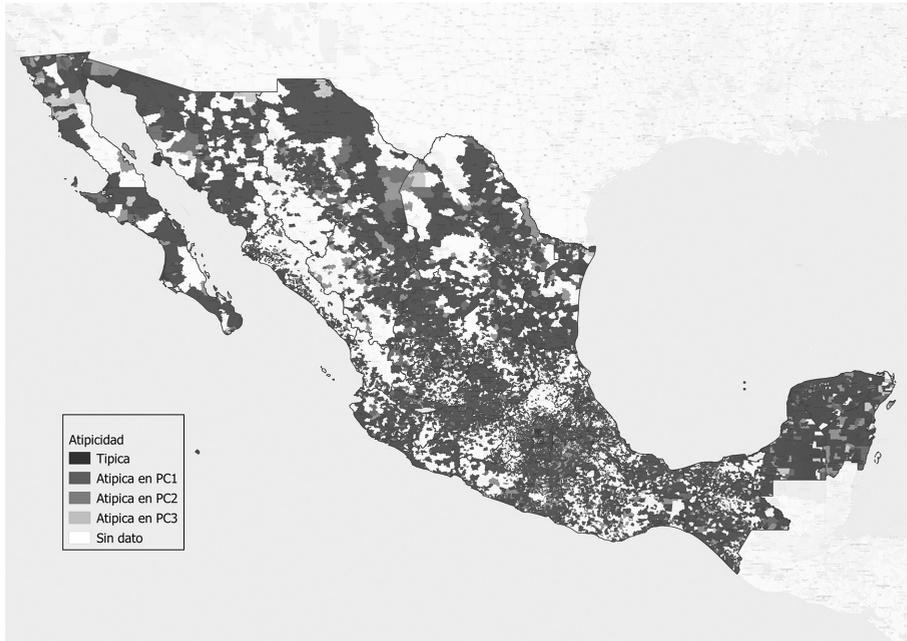
Anexo 2: Mapas

Figura 1. Secciones electorales y clasificación en la TCES



Fuente: Elaboración propia con qgis.

Figura 2. Secciones electorales atípicas



Fuente: Elaboración propia con QGIS.