

## **Insumo para el informe del componente cuantitativo del proyecto Tejidos Territoriales. Vínculos rurales-urbanos para el desarrollo inclusivo en Colombia**

### **1. Objetivos**

#### **Objetivo General:**

- Realizar un Análisis Exploratorio de Datos Espaciales (AEDE) de indicadores de movilidad, económicos y sociales con el fin proporcionar evidencia de la existencia de flujos rurales-urbanos entre municipios de los departamentos de Antioquia, Tolima, Valle del Cauca y Meta.

#### **Objetivos específicos:**

- Procesar y calcular los indicadores necesarios para la realización del análisis.
- Priorizar indicadores para su análisis individual.
- Realizar el Análisis de Componentes Principales (ACP) para calcular índices para las dimensiones de movilidad, económica y social.
- Utilizar el AEDE para estudiar los flujos rurales-urbanos de los indicadores priorizados y de los índices elaborados a través del ACP.
- Realizar un informe concreto sobre los resultados.

### **2. Justificación de la elección de las dimensiones y de las variables**

Las dimensiones y variables de análisis son el punto de partida del estudio. No obstante, estas decisiones deben estar en función tanto de la conceptual como de los métodos estadísticos que se utilizan en el análisis. Desde lo conceptual, las dimensiones deben estar en línea que consiste en que los vínculos rurales-urbanos hacen referencia a los flujos de personas, bienes, servicios, información, dinero y servicios ambientales. Adicionalmente, en cuanto a los métodos estadísticos, lo importante es llegar a resultados concretos e interpretables. El punto de partida con el cual se empalmó el proyecto estaba conformado por las siguientes dimensiones:

1. Flujos de personas.
2. Flujos de bienes y servicios.
3. Flujos de información.
4. Flujos de relacionamientos.
5. Flujos de gobernanza.
6. Controles.

Asimismo, la metodología recomendada por el equipo que estaba encargado anteriormente del componente cuantitativo para realizar el análisis de estas dimensiones e identificar flujos rurales-urbanos fue el Análisis de Componentes Principales (ACP). Empero, el nuevo equipo identificó dos aspectos importantes que hicieron reconsiderar las 6 dimensiones propuestas y el ACP como único medio para realizar el estudio. En primer lugar, la mayoría de dichas dimensiones incluían pocas variables. El ACP requiere un número mayor de variables por dimensión debido a que por

cuestiones estadísticas es probable que algunas se descarten. Del mismo modo, este método está pensado para reducir la dimensión de un gran número de variables para explicar un fenómeno latente. Por tanto, se planteó la posibilidad ante los demás miembros del equipo de reducir las dimensiones a tres de tal forma que cada una de estas incluyera un número considerable de variables para utilizar el método.

En segundo lugar, las variables ubicadas en cada una de estas dimensiones no medían directamente flujos como sus nombres lo indicaban. En este sentido, se propuso que todas las nuevas tres dimensiones no hicieran referencia directa a los flujos, sino que englobaran conceptos un poco más amplios y que los flujos se interpretaran como el resultado de las técnicas utilizadas en el análisis, dada la información disponible. El ACP será útil para generar índices agregados por cada dimensión y para realizar análisis de cluster con el fin de agrupar los municipios según su grado de similitud con respecto a los componentes principales de cada dimensión por departamento. Esta agrupación de municipios puede dar indicios sobre posibles flujos dentro de cada departamento debido a que, por ejemplo, si cuatro municipios del Tolima se desempeñan bien según los dos primeros componentes de la dimensión económica, es probable que existan vínculos entre estos municipios que jalonan la economía del departamento. Sin embargo, es difícil sostener esta afirmación debido a que se está dejando por fuera del análisis la dimensión espacial. Es más probable que estos vínculos existan si las agrupaciones toman en cuenta el territorio. Por ejemplo, si dichos municipios son Ibagué y sus vecinos geográficos, esta zona contigua que está conformada por estos cuatro municipios representa un área geográfica donde se intercambian bienes y donde existe un tejido territorial. Para lograr lo anterior, el nuevo equipo propone no solo hacer el ACP (que es insuficiente por las razones antes descritas), sino realizar un Análisis Exploratorio de Datos Espaciales (AEDE) con el fin de capturar la dimensión territorial y capturar la existencia de flujos a lo largo de la extensión geográfica de los departamentos de Antioquia, Tolima, Valle del Cauca y Meta.

Para finalizar esta sección, las tres dimensiones definidas en conjunto con todo el equipo fueron: conectividad, desempeño económico y condiciones sociales. La dimensión de conectividad está conformada por variables que miden directamente variables que miden flujos educativos, de información y de bienes y servicios. Las variables priorizadas en esta dimensión tienen que ver con cobertura de internet, cargas de bienes transportadas, operaciones comerciales, distancias de viaje, migración y migración violenta. Esto va en la línea con la dimensión de flujos rurales-urbanos mencionada al principio de esta sección. Igualmente, las variables priorizadas en la dimensión de desempeño económico se escogieron teniendo en cuenta algunos de los factores que según la literatura determinan el desempeño económico. En primer lugar, se priorizaron variables que miden directamente la actividad económica como valor agregado, variables del mercado laboral, entre otras. En segundo lugar, se priorizaron variables con corte institucional debido a que la literatura propone que la senda desarrollo de un país puede estar marcada por las diferencias en el desempeño institucional y se necesitan unas reglas de juego claras para garantizar el éxito económico (generación de recursos propios, ahorro, variables electorales, etc.) (○). En tercer lugar, se priorizaron variables que tienen que ver con la infraestructura porque existe evidencia de que este aspecto es uno de los principales determinantes del crecimiento económico (○). Algunas de estas variables fueron cobertura de acueducto, alcantarillado, etc. Igualmente, el ámbito

multidimensional es un componente importante para medir el desarrollo y la calidad de vida de los habitantes como lo explican los desarrollos teóricos de Sen, por lo cual se priorizaron variables que tienen que ver con ambiente, seguridad, salud y educación. Por último, las variables individuales para su análisis priorizadas por todo el equipo fueron:

- Intensidad de la migración bruta.
- Número de relaciones bidireccionales.
- Distancia al nodo de educación más cercano.
- Índice de migración violenta neta acumulado (2000-2016).
- Cobertura de internet rural.

## **2. Justificación de la elección de la metodología de componentes principales**

El ACP es una herramienta útil cuando se quiere reducir a dimensión de una matriz muy grande de variables y para hacer agrupamiento en función de los componentes principales que explican la mayor varianza de un proceso latente (desempeño económico, por ejemplo). El agrupamiento permite hacer una caracterización de las observaciones analizadas con el fin de identificar las que se encuentran mejor en términos de los componentes principales, los que se encuentran mal en términos de los componentes y los que se encuentran bien en uno y mal en otro. No obstante, estas agrupaciones no toman en cuenta la dimensión espacial, por lo cual se pueden crear conglomerados en donde las observaciones no sean vecinas desde la perspectiva territorial. Igualmente, los índices que resultan de la reducción de la dimensión de la matriz de variables serán útiles para capturar los fenómenos latentes de movilidad, desempeño económico y condiciones sociales. Estos índices serán un insumo muy relevante para el análisis espacial pues permitirán analizar las tres dimensiones elegidas de forma concreta con tres indicadores. En concreto:

- Este modelo nos arroja índices ordinales por cada dimensión: un indicador de 0-100, donde 0 es un resultado desfavorable y 100 es un resultado favorable.
- Los indicadores responden a las 3 categorías definidas que agrupan un grupo de variables (los de la base de datos).
- Posibilidades de análisis de los resultados del modelo: tablas de estadísticas descriptivas, gráficas de dispersión, aglomeraciones, índices agregados, otros.

## **4. Justificación de la elección de la metodología del análisis espacial exploratorio**

Se realizará un Análisis Socio-Espacial (ASE), que como lo explican Buzai y Baxendale (), se constituye como un subconjunto del Análisis Espacial (AE). En este tipo de análisis, según Peña (), los Sistemas de Información Geográfica (SIG) han jugado un papel protagónico, pues constituyen una herramienta que implica la unión de la información (ubicada espacialmente) y de herramientas informáticas para su análisis con objetivos concretos. No obstante, existen desarrollos incipientes en otro tipo de herramientas que se han consolidado como medios importantes en el AE que pueden llegar a utilizar las mismas técnicas que un SIG, un ejemplo es el programa estadístico R.

A partir de lo anterior, se puede entender que el Análisis Exploratorio de Datos Espaciales (AEDE) que se realizará en este trabajo entra en la categoría de ASE, con algunas variables económicas. El AEDE surge de la imposibilidad de las técnicas econométricas estándar de tratar la totalidad de los problemas surgidos cuando se trabaja con datos de naturaleza espacial: la heterogeneidad y dependencia espacial. Como mencionan Moreno y Vayá (), la primera tiene que ver con la variación en las relaciones en el espacio y la segunda hace referencia a la existencia de una relación funcional entre lo que pasa en un punto y lo que ocurre en otro lugar. Por tanto, al analizar la dependencia espacial de los distintas variables e índices, se están encontrando relaciones funcionales entre zonas geográficas que se pueden considerar como una aproximación a la identificación de tejidos territoriales. En este sentido, el AEDE se centra en el tratamiento de dichos efectos y se define como “el conjunto de técnicas que permiten describir distribuciones espaciales, identificar localizaciones atípicas (outliers espaciales), descubrir esquemas de asociación espacial (clusters espacial) y sugerir diferentes regímenes espaciales u otras formas de inestabilidad espacial” (). El AEDE que se realizará en este documento, tendrá entonces las siguientes etapas para cada departamento e indicador:

### 1. Elaboración de mapas coropléticos.

Los mapas coropléticos representan la forma más sencilla de visualizar la distribución geográfica de una variable. En general, las zonas de la región de estudio (municipios en nuestro caso) se somborean con base a los valores de la variable de estudio, lo cual permite tener una primera aproximación a la distribución geográfica de los datos y tener una intuición de si los datos se distribuyen de manera aleatoria o si existe cierto patrón geográfico. En este análisis se realizarán mapas de quintiles en donde las zonas más oscuras representarán las observaciones con valores más altos de las variables de interés.

### 2. Análisis de autocorrelación global y Diagrama de dispersión de Moran

En concreto, la autocorrelación global se mide a través de I de Moran Global (IMG), el cual es el producto cruzado de entre la variable de interés y su rezago espacial, con las variables en términos de desviaciones estándar respecto a su media. Para una observación  $i$ , esto se expresa como  $x_i - \bar{x}$ , donde  $\bar{x}$  es la media de la variable  $x$ .

$$I = \frac{N}{\sum_i \sum_j W_{ij}} \times \frac{\sum_i \sum_j W_{ij} (X_i - \bar{X})(X_j - \bar{X})}{\sum_i (X_i - \bar{X})^2}$$

$W_{ij}$  representa los elementos de la matriz de pesos espaciales,  $S_0 = \sum_i \sum_j W_{ij}$  es la suma de todos los pesos y  $N$  corresponde al número de observaciones. Los valores positivos (negativos) del índice indican autocorrelación espacial positiva (negativa), es decir, presencia de una concentración de valores similares (disímiles) de la variable en regiones vecinas. Para verificar la significancia estadística de los resultados, tal cual lo resaltan Buzai y Baxendale (), se recurre a un test de hipótesis, cuya hipótesis nula es la existencia de una configuración espacial aleatoria de los datos y la alternativa es que la configuración espacial de estos no es aleatoria.

Por otra parte, el diagrama de dispersión de Moran consiste en un gráfico en cuyo eje y se ubica el rezago espacial de la variable de interés y la variable original en el eje x. La pendiente de la

regresión lineal de esta gráfica es el IMG. Nosotros presentamos el gráfico en términos de desviaciones estándar respecto a la media. Con los pesos de la matriz estandarizados por filas, la expresión del IMG se reduce a

$$I = \frac{\sum_i \sum_j w_{ij} z_i \cdot z_j}{\sum_i z_i^2} = \frac{\sum_i (z_i \times \sum_j w_{ij} z_j)}{\sum_i z_i^2}.$$

Este gráfico ayuda a clasificar la naturaleza de la autocorrelación espacial en cuatro categorías. Como la gráfica está centrada en 0, todos los puntos a la derecha de la media tienen  $z_i > 0$  y los de la izquierda  $z_i < 0$ . Estos valores están etiquetados como Alto y Bajo, respectivamente. Igualmente, los valores de los rezagos espaciales se clasifican con el mismo criterio si su valor  $z_j$  está por encima o por debajo de la media como Alto y Bajo respectivamente. En este sentido, se forman cuatro cuadrantes de autocorrelación espacial con clasificaciones Alto-Alto, Bajo-Bajo, Alto-Bajo y Bajo-Alto. Esto representa una conexión entre la autocorrelación espacial global y local. La autocorrelación espacial local se explicará con más detalle posteriormente.

### **3. Análisis de Autocorrelación Espacial Local (AEL) - Local Indicator of Spatial Association (LISA).**

El análisis de autocorrelación global prueba la hipótesis nula de si la distribución espacial de los datos es aleatoria, pero no brinda información de la localización de los clusters, lo cual resulta clave para la identificación de relaciones funcionales o tejidos territoriales entre los municipios. Para solucionar este problema Anselin (1995) propuso el indicador local de asociación espacial, LISA, por sus siglas en inglés. Un LISA proporciona un estadístico para cada observación (municipio) con sus respectivo P-valor para evaluar significancia estadística. El LISA para la observación  $i$  se calcula de la siguiente forma:

$$I_i = \frac{\sum_j w_{ij} z_i z_j}{\sum_i z_i^2}.$$

La nomenclatura de la fórmula anterior es la misma que se utilizó para el I Moran Global. Un valor positivo de este índice indica que una unidad tiene unidades vecinas con valores de atributos similares (altos o bajos), mientras que un valor negativo demuestra la existencia de un caso atípico, es decir, una observación con unidades vecinas cuyos valores de atributos son diferentes.