



UNIVERSIDAD TECNOLÓGICA NACIONAL
FACULTAD REGIONAL TUCUMÁN

TESIS DE MAESTRÍA

**Aplicación de los Sistemas de Información
Geográficos en la redefinición de los circuitos
educativos de la Provincia de Tucumán**

Autor:
Ing. Rubén OLEA

Director:
Dr. Pedro ARAUJO

Co-Director:
Esp. Ing. Horacio
MADARIAGA

*Esta tesis cumple con los requisitos para el grado de
Magister en Ingeniería en Sistemas de Información*

en



Grupo de Investigación en Tecnologías Informáticas Avanzadas
Universidad Tecnológica Nacional - Facultad Regional Tucumán

12 de abril de 2021

Declaration of Authorship

I, Ing. Rubén OLEA, declare that this thesis titled, «Aplicación de los Sistemas de Información Geográficos en la redefinición de los circuitos educativos de la Provincia de Tucumán» and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:

Date:

«Thanks to my solid academic training, today I can write hundreds of words on virtually any topic without possessing a shred of information, which is how I got a good job in journalism.»

Dave Barry

UNIVERSIDAD TECNOLÓGICA NACIONAL
FACULTAD REGIONAL TUCUMÁN

Resumen

Universidad Tecnológica Nacional - Facultad Regional Tucumán

Aplicación de los Sistemas de Información Geográficos en la redefinición de los circuitos educativos de la Provincia de Tucumán

por Ing. Rubén OLEA

La evolución, a través del tiempo, del sistema educativo argentino hacía la universalización implicó llevar a cabo transformaciones en lo organizativo, hubo que generar sistemas de administración de grandes contingentes de alumnos, docentes y recursos necesarios. El sistema educativo está fuertemente ligado a lo territorial, ya que la distribución espacial de la población es heterogénea, y está estrechamente relacionada con factores de distinta índole. La provincia de Tucumán no está exenta de esta heterogeneidad en la distribución territorial de la población, ni sus organismos de gobierno, incluido el sistema educativo, lo están de tener que organizarse territorialmente según sus obligaciones y necesidades. En este contexto el Ministerio de Educación (ME) de la Provincia de Tucumán divide a su jurisdicción territorial en: a) Zonas para calcular la Bonificación por ubicación, b) Zonas de supervisión y c) Circuitos territoriales. Esta última división tiene una lógica basada en los límites administrativos. En este trabajo, a través de la utilización de Sistemas de Información Geográfica (SIG) y de distintos algoritmos de agrupamientos, proponemos como mejorar la división en circuitos territoriales.

Acknowledgements

The acknowledgments and the people to thank go here, don't forget to include your project advisor...

Índice general

Declaration of Authorship	III
Resumen	VII
Acknowledgements	IX
1. Introducción	1
1.1. Objetivos del Trabajo	3
1.2. Principales Contribuciones	3
1.3. Caso de Estudio	3
1.4. Plan del Documento	4
2. Marco teórico	5
2.1. Sistemas de Información Geográficos	5
2.1.1. Orígenes de los SIG	5
2.1.2. Definición y conceptos principales de los SIG	6
2.1.3. Tendencias de los SIG	8
2.1.4. Herramientas de los SIG	9
GRASS	9
GDAL	9
QUANTUM GIS	9
ARCGIS	9
MAPSERVER	10
POSTGIS	10
GEOSERVER	10
GEONODE	10
Resumen de las herramientas SIG	11
2.1.5. Los SIG en el análisis de datos	11
2.2. Análisis de Datos: agrupamientos	12
2.2.1. Definiciones y conceptos de agrupamiento	12
2.2.2. Métodos de agrupamiento	13
Algoritmo K-MEANS	14
Algoritmo PAM	14
Algoritmo CLARA	14
Algoritmo CLARANS	14
Algoritmo DBSCAN	15
Algoritmo KAMILA	15
2.2.3. Validación de agrupamientos de datos	15
APN (Average Proportion of Non-overlap)	16
AD (Average Distance)	16
ADM (Average distance between means)	16
FOM (Figure of merit)	16
Connectivity	17

	Dunn	17
	Silhouette	17
2.3.	Conclusiones	17
3.	Estado del arte	19
3.1.	Análisis espacial de datos	19
3.1.1.	Análisis espacial con técnicas estadísticas	21
3.1.2.	Análisis espacial con técnicas SIG	22
3.1.3.	Análisis espacial con técnicas de agrupamiento	22
3.2.	Regionalización	23
3.3.	Conclusiones	26
4.	Contexto actual	27
4.1.	Legislación y Situación Actual	27
4.1.1.	Resolución N° 217/1991 (Bonificación por ubicación)	28
4.1.2.	Metadatos	30
4.2.	Orígenes de los Datos a Utilizar	31
4.2.1.	Edificios Escolares	32
4.2.2.	Caminería	33
4.2.3.	Circuitos Educativos	34
4.2.4.	Elevación del Terreno	34
4.2.5.	Cuencas Hídricas	34
4.2.6.	Límites departamentales	35
4.2.7.	Límites municipales y comunales	36
4.2.8.	Ejidotes urbanos	36
4.2.9.	Asentamientos y edificios	37
4.2.10.	Recorrido colectivos	37
4.2.11.	Comisarías	38
4.2.12.	Efectores de salud	38
4.3.	Consideraciones sobre los datos	38
4.3.1.	Consideraciones generales	39
4.3.2.	Zona de alta montaña y difícil acceso	39
4.3.3.	Parámetro 1: Distancia	40
4.3.4.	Parámetro 2: Medios de Transporte Público u otros	41
4.3.5.	Parámetro 3: Hospitalidad de la Zona	41
4.3.6.	Parámetro 4: Ubicación	42
4.3.7.	Parámetro 5: Clima	42
4.3.8.	Parámetro 6: Salud	43
4.3.9.	Parámetro 7: Comunicaciones	43
4.3.10.	Parámetro 8: Seguridad	43
4.3.11.	Parámetro 9: Educación, Cultura y Recreación	44
4.3.12.	Parámetro 10: Economía Local	44
4.4.	Acondicionamiento de los datos obtenidos	44
4.4.1.	Distancia	44
4.4.2.	Distancias a rutas nacionales	45
4.4.3.	Distancias desde la escuela	45
4.4.4.	Cobertura del servicio de Red Eléctrica	45
4.4.5.	Montaña y llanura	45
4.4.6.	Almacenamiento de los datos	45
4.5.	Conclusiones	46

5. Resultados obtenidos	47
5.1. Determinación del Puntaje para el Cálculo de la Bonificación por Ubicación	47
Ejemplo	48
5.2. Métodos basados en densidad	50
5.2.1. DBSCAN	50
5.3. Métodos de particionado	51
5.3.1. K-MEANS	51
5.3.2. PAM	54
5.3.3. CLARA	54
5.3.4. KAMILA	56
5.4. Comparación de resultados	58
5.5. Miscelaneas	59
5.6. Conclusiones	60
6. Conclusiones y Trabajos Futuros	63
6.1. Conclusiones Generales	63
6.2. Perspectiva y Trabajos Futuros	64

Índice de figuras

2.1. Mapa de John Snow que representa la ubicación de los decesos por cólera.	5
2.2. Capa temática o layer (Fuente GAO). En la imagen se muestra como los datos pueden componer un mapa.	7
2.3. Tabla métodos de agrupamiento.	15
3.1. Taxonomía de los métodos para resolver problemas de regionalización [46].	24
4.1. Comparación de los circuitos territoriales con los límites departamentales.	28
4.2. Localización de los edificios escolares.	33
4.3. Mapa de la caminería de la Provincia de Tucumán.	34
4.4. Relieve de la provincia	35
4.5. Mapa de las cuencas hídricas de la Provincia de Tucumán.	35
4.6. Departamentos de la Provincia de Tucumán.	36
4.7. Municipios y comunas de la Provincia de Tucumán.	36
4.8. Asentamientos y edificios.	37
4.9. Comisarías.	38
4.10. Efectores de salud.	39
4.11. En rojo la ubicación de escuelas, en amarillo el faltante de ubicar (un CAM y una comisaría) en los datos del IGN.	40
4.12. Error de escala.	40
4.13. Datos obtenidos a partir del DEM.	42
5.1. Vinculación entre las distintas variables utilizadas.	47
5.2. Se calcularán los parámetros correspondientes a la coordenada indicada.	48
5.3. Resultados del método DBSCAN con la totalidad de datos.	51
5.4. Resultados del método k-means con la totalidad de datos.	52
5.5. Resultados del método k-means donde se observa un sesgo dado por la variable departamento.	53
5.6. Resultados del método k-means con los datos puntaje total, elevacion, distancia, latitud y longitud.	53
5.7. Resultados del método PAM.	55
5.8. Resultados del método PAM.	55
5.9. Resultados del método CLARA con todas las variables.	56
5.10. Resultados del método CLARA con variables continuas.	57
5.11. Resultados del método Kamila, para 5, 6, 7, 8, 9, y 10 agrupamientos.	57
5.12. Resultados del método kamila sobre exactamente el mismo conjunto de datos, para obtener 5 agrupamientos.	58

Índice de cuadros

3.1. Resumen de trabajos revisados sobre Análisis Espacial de Datos, tomando en cuenta algoritmos empleados, objetivos propuestos y metodología.	21
3.2. Resumen de trabajos de regionalización.	25
4.1. Puntajes máximos y mínimos posibles de ser asignados.	30
5.1. Ejemplo de resultado obtenido para una coordenada en particular. . .	50
5.2. Comparación de métricas estadísticas aplicadas a los resultados. . . .	60

List of Abbreviations

AD	Average Distance
ADM	Average Distance Between Means
APN	Average Proportion of Non-Overlap
BD	Base de Datos
BDE	Base de Datos Espacial
CS	Servicio de Catálogo
CAC	Centro de Acción Comunitaria
CAM	Centro de Salud Municipal
CAPS	Centro de Atención Primaria de Salud
CIC	Centro Integrador Comunitario
CoNaIISI	Congreso Nacional de Ingeniería Informática/ Sistemas de Información
COVID-19	Coronavirus Disease 2019
CS	Catalogue Service
DEM	Digital Elevation Model
ESS	Error Sum of Squares
FOM	Figure Of Merit
GNU GPL	GNU General Public License
GPS	Global Positioning System
GSMT	Gran San Miguel de Tucumán
IDE	Infraestructura de Datos Espaciales
IDET	Infraestructura de Datos Espaciales de Tucumán
IGN	Instituto de Geográfico Nacional
INDEC	Instituto Nacional de Estadística y Censos
INTA	Instituto Nacional de Tecnología Agropecuaria
JAIIO	Jornadas Argentinas de Informática e Investigación Operativa
ME	Ministerio de Educación
MIMO	Map In – Map Out
MSC	Ministerio de Seguridad Ciudadana
MST	Ministerio de Salud de Tucumán
OGC	Open Geospatial Consortium
OSM	OpenStreetMap
RIDES	Red de Información para el DESarrollo Productivo
SIG	Sistema de Información Geográfica
SIGC	Sistema de Información Geográfica de Canadá
SMT	San Miguel de Tucumán
SR	Sistema de Referencia
SRTM	Suttle Radar Topography Mission
TIG	Tecnologías de Información Geográfica
USGS	United States Geological Survey
WCS	Web Coverage Service
WFS	Web Feature Servide
WMS	Web Map Service
WS	Web Service

For/Dedicated to/To my...

Capítulo 1

Introducción

La República Argentina, a lo largo de su historia y especialmente en el comienzo de este nuevo milenio, avanzó en la definición de normativas que garanticen y mejoren el derecho a la educación de toda su población. La Ley de Educación Común N° 1.420, sancionada en 1.884 (primera ley de educación de la República Argentina) definió la escuela primaria como obligatoria y gratuita. La evolución de la matrícula primaria creció junto a la población a lo largo del tiempo. En cambio, la matrícula secundaria crece a través de saltos respecto del tamaño de la población, dándose el primero durante el peronismo, el segundo a partir del año 1.983 con el retorno de la democracia. Otro salto se da con la sanción de la Ley Federal de Educación, N° 24.195 (año 1.993), que establece la obligatoriedad de 10 años de escolaridad. En el año 2.006, la sanción de la Ley de Educación Nacional N° 26.206, extiende la obligatoriedad de la educación a toda la secundaria, y establece la universalización de la sala de cuatro años del nivel inicial [6]. El decreto N° 1.602/09 crea la Asignación Universal por Hijo, el cual *“ha fomentado el acceso a los servicios de salud y la escolarización de sus beneficiarios, especialmente en el nivel secundario. Varios autores destacan que esta política ha significado un aumento considerable de la matrícula escolar”* [47].

Esta evolución hacia la universalización implica profundas transformaciones. *“Para la política educativa esto supone una disminución sensible de la profundidad en favor de la extensión”* [48]. Sobre todo, estos cambios deben ser llevados a cabo desde lo organizativo a nivel macro, ya que habrá que generar sistemas de administración de inmensos contingentes de alumnos, docentes y escuelas. Esto implica tanto el crecimiento de la inversión (logrado en el año 2005 con la Ley de Financiamiento Educativo N° 26.075), como la capacidad del planeamiento y organizacional del sistema educativo.

A pesar de los resultados obtenidos para lograr la universalización de la educación, el sistema educativo se encuentra en un proceso de crisis que afecta a muchos de los aspectos que durante décadas no fueron cuestionados [49]. La educación debe adaptarse a la situación de crisis, donde los beneficiarios de dicho derecho necesitan conocimiento y herramientas. *“Las crisis económicas involucran dinámicas de exclusión, concentración y presión sobre los recursos”* [7]. A su vez, la educación está fuertemente ligada a lo territorial, ya que su tarea en sus relaciones con el territorio, es la de movilizar todo su talento humano, social e intelectual para que el sujeto-ciudadano entre al mundo y se sostenga en él [50]. La permanencia de los alumnos en el sistema educativo está fuertemente vinculada a diversos factores, entre los que podemos identificar las condiciones socio-económicas del territorio en el que habitan, las características del mercado laboral en el que están insertos los adultos, el acceso a los servicios básicos como salud e infraestructura, acceso a las tecnologías de la comunicación, entre otras. La distribución espacial de la población es heterogénea, y está estrechamente relacionada con factores de carácter geográfico, económico, social, político, ambiental, cultural e histórico [51]. Por ello, podría decirse que la

concentración de población en determinadas regiones no es más que una reacción a una mayor disponibilidad de recursos, infraestructura y servicios, así como también la dispersión de población responde a la falta de los factores indicados.

El gobierno del sistema educativo se ejerce a través de los organismos centrales de decisión, y también mediante el accionar de niveles intermedios que se ubican entre las instancias de toma de decisiones y las escuelas, y que involucran múltiples actores. Es así que, la descentralización se constituye en la estrategia política que permitirá la representación y presencia de un Estado cercano a las realidades locales. La descentralización, entonces, trae consigo connotaciones en términos de toma de decisiones oportunas, adecuadas y contextualizadas.

A la descentralización educativa, como a toda política pública, le toca desafiar una complejidad de problemas, entre los que se destaca la dimensión territorial [56]. Esta dimensión adquiere relevancia por la asignación de valores que las sociedades locales dan al contenido de la política descentralizadora. El potencial de descentralización supone la existencia de un sujeto capaz de asumir la gestión de intereses colectivos, dotado de personalidad socio-cultural y político-administrativa y la transferencia a este sujeto de competencias y recursos (de distinta índole) que ahora puede no tener y que deberá gestionar autónomamente [56].

La Ley de Educación N° 26.206 [99] afirma que el Estado debe garantizar las condiciones para que todos los alumnos logren aprendizajes de calidad independientemente de su origen social, radicación geográfica, género o identidad cultural. Esto también puede inferirse al docente en general, puesto que es responsable de lograr calidad en el aprendizaje de sus alumnos. Entonces es importante formular la pregunta acerca de cómo impactará el proceso descentralizador en la cobertura, calidad y eficiencia del sistema educativo en su cometido de garantizar la igualdad de oportunidades para la ciudadanía [56].

Los circuitos educativos colaboran en la descentralización educativa, pero por sí mismas no delimitan zonas de responsabilidad, como lo haría un límite administrativo. Esta organización territorial promueve la articulación de recursos, capacidades y por sobre todo potencialidades, posibilita esquemas de cogestión entre los distintos niveles de ejecución del sistema educativo y otros organismos de gestión [34]. La conformación de los circuitos tuvo como fundamento evidenciar mayor calidad, integralidad, flexibilidad y relevancia de la oferta educativa [34].

Es importante formular la pregunta acerca de cómo impactará el proceso descentralizador en la cobertura, calidad y eficiencia del sistema educativo en su cometido de garantizar la igualdad de oportunidades para la ciudadanía [56].

La situación impuesta por la pandemia del COVID-19 durante el año 2020, condicionó al sistema educativo al desafío de adoptar rápidamente medios virtuales como herramientas pedagógicas para el desarrollo de los programas educativos, generando incertidumbre y ansiedad en la comunidad educativa [103]. Sin embargo la no obligatoriedad del trasladarse al lugar de trabajo/estudio, puede interpretarse como un beneficio que en el tiempo tendrá que anularse para volver a la normalidad.

Esta normalidad tendrá nuevamente como factor predominante la distancia a recorrer y los tiempos de traslado para cumplir con la obligación de impartir educación o el derecho de recibirla. Siendo necesario el repensar la organización territorial para mejorar los beneficios de la descentralización.

1.1. Objetivos del Trabajo

El objetivo general de esta tesis consiste en el estudio, definición y aplicación de nuevos circuitos territoriales educativos en la Provincia de Tucumán tomando en cuenta los diferentes factores y características de la provincia como la división política, distribución poblacional, accidentes geográficos, etc.

A continuación se describirá brevemente los objetivos específicos de la tesis:

- Recolectar datos de la ubicación de los establecimientos y la conformación de las distintas zonas de supervisión.
- Generar información espacial a partir de datos existentes, georeferenciando los datos existentes, o generando layers a partir de la información alfanumérica utilizando metodologías GIS.
- Realizar un estudio del estado del arte respecto de soluciones al problema en estudio, utilizando metodologías SIG.
- Plantear distintas metodologías SIG que permitan lograr el objetivo general.
- Evaluar y seleccionar técnicas de agrupamientos que en conjunto con las metodologías SIG permitan lograr el objetivo general.
- Encontrar la solución más adecuada para el problema planteado.
- Zonificar la provincia, según la estructura organizacional del ME, adecuando a las políticas territoriales definidas.

1.2. Principales Contribuciones

En el marco de desarrollo de esta Tesis de Maestría, se trabajó principalmente en la redefinición de los circuitos educativos del Ministerio de la Provincia de Tucumán. Esta tesis representa una continuación a otros trabajos que el alumno intervino en su actividad diaria en el Ministerio de Educación.

Paralelamente a este desarrollo, se ha trabajado en numerosas temáticas referidas a la educación con datos que se encuentran libremente disponibles en el Ministerio de Educación de Tucumán, Ministerio de Educación de la Nación y otros organismos gubernamentales de la esfera nacional y provincial.

Parte de los resultados obtenidos en esta tesis han sido comunicados en el congreso 49 Jornadas Argentina de Informática (49 JAIIO) donde se presentó el artículo titulado "*Aplicación de los Sistemas de Información Geográfica (SIG) en el análisis de la bonificación por ubicación*". Posteriormente se realizó la presentación de un nuevo artículo denominado "*Propuesta de mejora de los circuitos educativos utilizando Sistemas de Información Geográfica (SIG)*", en el 8vo Congreso Nacional de Ingeniería Informática/Sistemas de Información (CoNaIISI).

1.3. Caso de Estudio

El problema de abordar lo territorial, que encara esta tesis de maestría, es uno de los múltiples aspectos que afronta el sistema educativo de Tucumán.

En la actualidad, el Ministerio de Educación de Tucumán cuenta con una división territorial denominada circuitos educativos, la cual definió en base a los límites

departamentales existentes, a nuestro criterio de una forma que conjuga regiones de la provincia dispares (unos de sus circuitos contiene una territorialidad muy dispar).

El desafío fue realizar una nueva redefinición de dichos circuitos, con nuevos y sólidos criterios a partir de información que el Ministerio de Educación de Tucumán y otros organismos gubernamentales publican para la ciudadanía en general.

1.4. Plan del Documento

El presente documento está estructurado de la siguiente manera:

- El Capítulo 2 presenta una introducción a los SIG, haciendo foco en el análisis de datos con este tipo de técnicas. A su vez, se da una introducción al conjunto de herramientas utilizadas para el correcto desarrollo de esta tesis.
- En el Capítulo 3 se describe algunos de los trabajos más relevantes sobre regionalización que existen en la literatura científica.
- El Capítulo 4 describe el contexto actual de la Provincia de Tucumán y la división territorial en circuitos educativos. En el mismo, se hace referencia a la evolución del sistema educativo producto de la aplicación de las políticas nacionales en el ámbito de la educación
- En el Capítulo 5 se presenta la contribución principal de la tesis: la redefinición de los circuitos educativos de la Provincia de Tucumán. Para lograr esto se describen los métodos utilizados para la nueva regionalización de dichos circuitos, las ventajas y desventajas de cada uno. Se identifican y seleccionan las mejores soluciones para determinar como será la nueva distribución de los circuitos educativos y su justificación.
- Por último, en el Capítulo 6 se realiza una recapitulación de las principales contribuciones de la tesis, resaltando los aspectos más destacados de la misma. A su vez, se elabora la conclusión general del trabajo y se plantean los temas que constituirán los puntos de partida para el desarrollo de futuros trabajos.

Capítulo 2

Marco teórico

2.1. Sistemas de Información Geográficos

La geografía tiene un rol importante en las sociedades humanas, incide plenamente en la vida humana. Junto con la cartografía tienen como punto en común el espacio geográfico, mientras la primera capta el espacio en su relación causa-efecto, la segunda registra de manera fiel y exacta los elementos del terreno [104]. La cartografía no fué ajena al desarrollo tecnológico, su evolución se impulsó fuertemente gracias a los sistemas computacionales, que permitieron masificar la producción y uso de cartografía, como así también la captura y generación de datos geográficos [110].

2.1.1. Orígenes de los SIG

En 1.832 el geógrafo francés Charles Picquet creó un mapa que representaba un brote de cólera en 48 distritos de París. Este mapa fue una versión temprana de un mapa de calor. John Snow, en Londres en 1.854, adoptó el mismo principio para registrar en un mapa (Fig. 2.1), a través de puntos que representaban, la ubicación de los lugares donde vivían las víctimas del cólera. Identificó así, que las personas que bebían agua proveniente de una fuente de agua cercana, tuvieron una alta tasa de mortalidad. Si bien la cartografía ya existía, el mapa de Snow fué el primero en no sólo representar la realidad, sino que permitió analizar fenómenos espaciales dependientes.

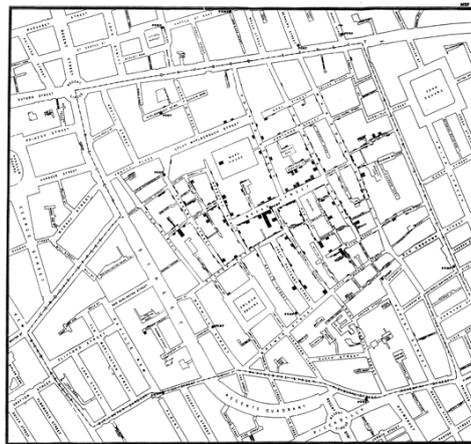


FIGURA 2.1: Mapa de John Snow que representa la ubicación de los decesos por cólera.

En 1959, Waldo Tobler, al aplicar los ordenadores al campo de la cartografía, publicó el primer mapa hecho en computadora basado en su algoritmo MIMO (Map In-Map Out) [105]. A comienzos de los años 60, se observa el desarrollo de las primeras aplicaciones cartográficas para computadoras. El Sistema de Información Geográfica de Canadá (SIGC)¹ fue el primer SIG en el mundo similar a tal y como los conocemos hoy en día, sin embargo, este sistema no fué comercializado. La aparición de estos programas no solo implica la creación de una herramienta nueva, sino también el desarrollo de técnicas nuevas que hasta entonces no habían sido necesarias. La más importante de ellas es la codificación y almacenamiento de la información geográfica.

En los años 80, emergen las primeras aplicaciones comerciales de SIG, tomando como base el SIGC. En 1.984 aparece GRASS² y en 1.999 en su versión 5.0 fué lanzada bajo licencia GNU General Public License (GNU GPL), siendo aún en la actualidad el referente dentro de su área. Es en esta década que se comienza a observar inconvenientes al momento de compartir datos espaciales, debido a una falta de estandarización [109]. Otro gran impulso en el uso de los SIG se dió a partir de la accesibilidad de imágenes satelitales y el acceso al Sistema de Posicionamiento Global (Global Positioning System ó GPS). El GPS es un sistema que permite determinar en toda superficie de la Tierra la posición de cualquier objeto, gracias a una constelación de satélites de uso militar exclusivo hasta ese momento.

A fines del siglo XX, se comienzan a migrar los conceptos de visualización de datos SIG a Internet. En los años 90 aparecen los primeros servidores de mapas, como Xerox PARC Map viewer y/o mapserver, siendo este último uno de los principales servidores en la actualidad, con lo que se facilita el poder compartir la información espacial. Estas iniciativas despejan el camino para un nuevo concepto. En 1993 el U.S. National Research Council acuña el término Infraestructura de Datos Espaciales (IDE), en 1994 el gobierno estadounidense crea la National Spatial Data Infrastructure (NSDI) [132]. Estas iniciativas sumadas a la creciente necesidad de estandarizar los formatos de datos y normas de transferencia marcaron el nacimiento de la Open Geospatial Consortium (OGC)³ [109]. La OGC es una organización sin fines de lucro que define estándares abiertos para la comunidad geoespacial global. Estas acciones sembraron la semilla de la democratización de los datos espaciales.

En los últimos años, hubo una expansión en el número de desarrollos de software SIG de código libre, los cuales, a diferencia del software comercial, suelen abarcar una gama más amplia de sistemas operativos. En nuestros días, aplicaciones como Google Maps o Google Earth y la multitud de aplicaciones con interfaces Web basadas en ellas, permiten acceder a información geográfica de toda índole, los SIG han pasado a ser elementos de consumo y estar presentes en nuestra vida diaria.

2.1.2. Definición y conceptos principales de los SIG

En un sentido genérico, los SIG son herramientas que permiten a los usuarios crear consultas interactivas, analizar la información espacial y presentar los resultados de todas estas operaciones. En el sentido más estricto, es cualquier sistema de información capaz de integrar, almacenar, editar, analizar, compartir y mostrar la información geográficamente referenciada. Los SIG como herramienta, permiten resolver problemáticas con una fuerte componente espacial lo que nos ayuda a lograr una mayor eficiencia en el estudio de las problemáticas vinculadas a lo territorial.

¹<https://canadiangis.com/>

²<https://grass.osgeo.org/>

³<http://www.opengeospatial.org>

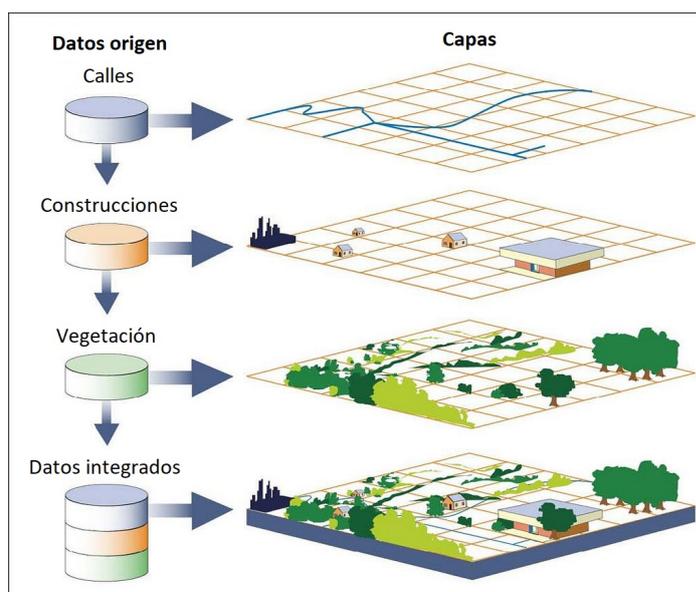


FIGURA 2.2: Capa temática o layer (Fuente GAO). En la imagen se muestra como los datos pueden componer un mapa.

Los SIG permiten la integración de números ilimitados de variables, representando cada uno de ellas un tema espacial (red de caminos, hidrografía, ocupación del suelo, etc.). Cada una de estas variables recibe el nombre de capa temática (layer), pudiéndose utilizar diferentes algoritmos para llevar a cabo geoprosesos entre ellas. La Figura 2.2 muestra el concepto de capa temática o layer en SIG. Un SIG posee una estructura compuesta por:

1. *Datos*: la materia prima con la que podemos construir productos de información o mapas que nos ayuden a hacer nuestros análisis y tomar las decisiones en nuestra organización. De ellos dependen, por ejemplo, los procesos que podremos o no realizar con los mismos y en general todo cuanto podemos esperar de ellos;
2. *Análisis*: métodos y procesos enfocados al análisis de los datos: el análisis es una de las razones fundamentales que llevaron al desarrollo de los SIG. Un SIG siempre incorpora una serie de formulaciones que permiten la obtención de resultados y el análisis de los datos espaciales. Las formulaciones representan procesos que pueden ser sencillos o sumamente complejos, algunos de ellos se aplican en otros campos. Su origen no siempre es el ámbito puro de la geografía, por ejemplo, la estadística aporta a los SIG muchas de sus ideas constituyendo un nuevo conjunto de procesos de análisis;
3. *Visualización*: visualizaciones, métodos y fundamentos relacionados con la representación de los datos: no es solo una forma más de trabajar con los datos, es la principal. La información geográfica tiene una inherente naturaleza visual, ya que el ser humano entiende al espacio en forma gráfica;
4. *Tecnología SIG*: se incluye aquí tanto el hardware como el software, los cuales conforman un binomio donde encontraremos distintas alternativas que se enriquece con la evolución de la tecnología;
5. *Factor organizativo*: engloba los elementos relativos a la coordinación entre personas, datos y tecnología, o la comunicación entre ellos, entre otros aspectos.

Como ejemplo, respecto de los usuarios tenemos distintos perfiles, desde data entrysts, analistas, tomadores de decisión, y con la aparición de la WEB 2.0 y el bajo costo de las unidades GPS, el SIG ha llegado a usuarios no especializados. Resulta evidente el gestionar correctamente la complejidad del sistema SIG [8].

2.1.3. Tendencias de los SIG

El uso masivo de Internet y de las tecnologías de comunicación ofrecen nuevas perspectivas para los SIG mediante el acceso a las bases de datos geográficas distribuidas [9]. Sobre la base de arquitecturas orientadas a servicios, los servicios web (Web Service ó WS) permiten intercambiar información entre aplicaciones y procesar datos geoespaciales; siendo los WS y los estándares de la OGC son adecuados para este propósito [55]. Existen distintos tipos de WS espaciales, podemos citar entre otros: - Servicio Web de Mapas (Web Map Service ó WMS), que permite la descarga de un mapa en un formato gráfico, es decir una imagen, asociada a su información espacial [30]; - el Servicio Web de Características (Web Feature Service ó WFS): permite la descarga de layers de información geográfica vectorial completa, es decir, la descarga de su geometría y de su tabla de atributos asociada [31]; - Servicio Web de Coberturas (Web Coverage Service ó WCS): que permite realizar peticiones de coberturas geográficas (imágenes satelitales o fotografías aéreas) [29]; - Servicio de Catálogo (Catalogue Service ó CS): permite la publicación de la información descriptiva (metadatos) de los datos, servicios y objetos que son compartidos [62].

Los metadatos describen el contenido, calidad, condición, entre otros aspectos de los datos geográficos. Permiten localizar y entender los datos geográficos publicados, ya que nos brindan información sobre quienes produjeron el dato, los atributos que poseen, la fecha del dato, el sistema de referencia de los datos, etc [59]. Uno de los datos más importantes dentro de los metadatos es el Sistema de Referencia (SR) de los datos. Las proyecciones cartográficas intentan representar la superficie de la tierra, o una parte de ella, en una superficie plana (por ejemplo, el papel o la pantalla de la PC), un SR define como el mapa bidimensional se relaciona con la figura en tres dimensiones que intenta reproducir. Como ya dijimos una de las funciones de los SIG es recolectar datos, estos se generarán bajo diferentes métodos, se almacenarán de distintas formas; por lo general si los datos no proceden de la misma fuente es muy probable que tengan distintos SR [10, 11].

La cada vez mayor disponibilidad de datos espaciales y el desarrollo de los SIG ha incrementado el uso de los mismos, en distintos aspectos del trabajo cotidiano y de la investigación por parte de los usuarios medios de los SIG. Este creciente uso ha sido exponencial en los últimos tiempos, y tal fué su magnitud que aparecieron nuevos paradigmas en este campo, que hacen no ya hablar de SIG en forma genérica sino de las Tecnologías de Información Geográfica (TIG) [112]. Las TIG conforman un conjunto de tecnologías cuyo núcleo está formado por la Teledetección, los SIG, los sistemas de Posicionamiento Global [114]. Este crecimiento en el uso de estas tecnologías está produciendo la democratización de las TIG [113], este fenómeno se acompaña con la generación y la demanda de geoinformación [111]. Frente a este fenómeno que no es propio de las TIG sino del conjunto de las nuevas tecnologías, apareció el Big Data que ha supuesto uno de los grandes hitos de la sociedad digital [111]. El término Big Data referencia a la masiva producción de datos por medio de la utilización de sensores y dispositivos; vivimos una revolución de los datos, los cuales adquieren un valor creciente para las organizaciones y la sociedad en su conjunto [115]. El uso de Big Data se está extendiendo en diversos campos, ya que permite el análisis de tendencias y patrones para la toma de decisiones.

2.1.4. Herramientas de los SIG

Waldo Tobler publicó el primer mapa realizado en computadora en 1959, al aplicar los ordenadores al campo de la cartografía. La aparición de programas como SIGC no sólo implicaron la aparición de una nueva herramienta, sino también el desarrollo de técnicas nuevas que no habían sido necesarias hasta entonces. Se puede decir que su naturaleza de representar la realidad utilizando la tecnología disponible hace a los SIG útiles y en permanente evolución, estando siempre entre las nuevas tecnologías [116]. En esta sección presentaremos algunas de las herramientas SIG más populares:

GRASS

GRASS⁴ (Geographical Resources Analysis Support System) fue desarrollado por el Cuerpo de Ingenieros del Ejército de los EE.UU., cuando el Laboratorio de Investigación de Ingeniería de Construcción comenzó a explorar el uso de SIG. Su desarrollo comenzó en 1982 y fue lanzado al público en general en 1984. En 1999, con su versión 5.0 se lanzó bajo GNU GPL. Es un software con una fuerte orientación hacia la investigación [118]. Su punto más débil es que estuvo diseñado para entornos UNIX, sin embargo al contar con más años de desarrollo que otros productos cuenta con un elevado número de herramientas y utilidades. Originalmente estuvo orientado al tratamiento de imágenes ráster [119]. GRASS tiene la capacidad de poder ser utilizado como complemento por parte de otro software.

GDAL

Debido a la diversidad de formatos ráster de datos espaciales comenzó a dificultarse su procesamiento homogéneo. Con el objetivo de lograr una accesibilidad genérica a estos formatos, Frank Warmendan en 1998, propone la librería GDAL. La principal característica de este desarrollo es que presenta un único modelo de datos abstracto que logra manipular la mayoría de formatos de datos ráster. Además GDAL fue desarrollado con el concepto de extensibilidad, es decir permite incorporar la posibilidad de manipular nuevos formatos a su biblioteca. Contiene, además, un conjunto de comandos para el procesamiento de rásters [131].

QUANTUM GIS

QUANTUM GIS o simplemente QGIS es software libre y de código abierto, desarrollado en C++. Fue uno de los ocho primeros proyectos de OSGeo. Está orientado al uso de complementos que pueden ser desarrollados en C++ o en Python. Esta capacidad se aprovecha para ser utilizado como interfaz gráfica de productos como GRASS, GDAL, SAGA y R entre otros. Este desarrollo tiene versiones que pueden funcionar en la mayoría de los sistemas operativos que existen. Desde sus comienzos fue adquiriendo cada vez mayor popularidad en el mundo de los SIG.

ARCGIS

ESRI, fundada en 1969, es una empresa líder en el mercado de los SIG. En el año 1981 lanzó al mercado su producto ArcInfo; a fines de los '90 se realizó una revisión total del producto de donde nace ARGIS, que está compuesto por diversos módulos

⁴<https://grass.osgeo.org/>

siendo básicos algunos y otros para aplicaciones muy especializadas [118]. La popularidad de ArcInfo y su versión ArcView dejaron, entre otras cosas, como estándar para compartir información espacial el formato SHAPEFILE. Una de las mayores ventajas de los productos de ESRI es el contar con soporte técnico en forma continua con rápidas respuestas, además las aplicaciones se van ajustando en función de las necesidades de los usuarios. En cambio, su mayor desventaja es el elevado costo de las licencias [120].

MAPSERVER

MAPSERVER es un software de código abierto, ampliamente utilizado, que permite visualizar cartografía vía internet [121], esto le permite funcionar como un servidor de mapas. Fue desarrollado en sus orígenes en la Universidad de Minnesota en colaboración con la NASA y el Dpto de Recursos Naturales de Minnesota [121]. Su mayor desventaja es la de no contar con una interfaz gráfica para su utilización por parte de los usuarios [122], en cambio su mayor ventaja es su performance al trabajar con imágenes ráster. Es un software multiplataforma, muchas aplicaciones WEB lo utilizan como motor para la publicación de cartografía.

POSTGIS

La extensión PostGIS permite al motor de Base de Datos (BD) PostgreSQL manejar datos espaciales. Esto es debido a que PostGIS incorpora tipos de datos espaciales, junto a una biblioteca de funciones que pueden operar sobre estos tipos de datos [123]. Esta extensión fue desarrollada por la empresa canadiense Refraction Research y nació como Open Source. Su mayor ventaja es su veteranía, demostrando a través del tiempo su eficiencia. En año 2006 fue certificado por la OGC.

GEOSERVER

GeoServer es un servidor Open Source de datos espaciales, cumple con los estándares abiertos para proveer de una variedad de fuentes de datos [124]. Su primera versión fue lanzada en el año 2001, está desarrollado en Java y ofrece una gran flexibilidad a los usuarios para crear y compartir información geográfica. Su mayor ventaja es la de poseer una interfaz que le permite a cualquier usuario poder administrar la herramienta.

GEONODE

GEONODE es un gestor de contenidos geoespaciales, desarrollado bajo el concepto de software libre. Es una plataforma para la administración y la publicación de datos espaciales. Permite desarrollar SIG y lo que sería su mayor potencial: el permitir desplegar IDEs. Entre sus componentes podemos citar Python, PostgreSQL, PostGIS, GeoServer, Django, entre otros. Entre sus ventajas podemos indicar es la sencillez de su interface; es un proyecto colaborativo, además tiene una documentación bastante completa. Entre sus desventajas podemos indicar que es tedioso para instalar en el entorno Windows.

Resumen de las herramientas SIG

En esta sección se presentaron sólo algunas de las herramientas más populares de los SIG. Existen muchas otras en el mercado, el cual está en constante evolución por lo que aparecen nuevas versiones de las herramientas existentes, como así también nuevos desarrollos, que tienden a integrarse con los ya existentes complementándose entre sí para brindar mayores prestaciones a los usuarios. Se realizó un muy breve recorrido por las herramientas de SIG existentes, mencionando tanto Open Source como software propietario, herramientas de escritorio como también utilizadas como servidores de mapas, herramientas que complementan otras como GRASS y POSTGIS.

2.1.5. Los SIG en el análisis de datos

Los SIG son herramientas que permiten capturar, almacenar, manipular, visualizar, analizar y compartir información geográfica. Nos permiten alcanzar una mayor eficiencia en el estudio de las problemáticas vinculadas a lo territorial.

Los SIG a través del análisis de datos permiten descubrir estructuras espaciales y relaciones entre los datos, así como el modelar fenómenos geográficos. Los resultados reflejan la naturaleza y calidad de los datos, como también la pertinencia de los métodos y funciones aplicadas. Las tareas y transformaciones que se llevan a cabo en el análisis espacial precisan datos estructurados, programas con las funciones apropiadas y conocimientos sobre la naturaleza del problema para definir los métodos de análisis. El proceso convierte los datos en información útil para conocer un problema determinado. Es evidente que los resultados del análisis espacial añaden valor y, sobre todo, información y conocimiento a los datos geográficos.

El análisis es una función específica de los SIG y es su elemento característico. Se utilizan procedimientos de análisis territorial para la obtención de resultados. En realidad son modelos resultantes del tratamiento de la información, los cuales están condicionados por las hipótesis contempladas y la calidad de los datos.

El análisis espacial refiere a aquellas ramas de análisis de datos en donde la referencia geográfica de los elementos contiene información importante. En muchas áreas, los índices que distinguen los casos observados pueden intercambiarse ya que la información pertinente para la comprensión de las variables está contenida en las observaciones, mientras que la no relevante está contenida en los índices. En cambio en los datos espaciales la información contenida en los índices (de localización y tiempo) puede ser de crucial importancia [130].

De las muchas definiciones del análisis espacial se puede indicar que representa un conjunto de técnicas y modelos que utilizan explícitamente la referencia espacial de los datos. Se requiere establecer supuestos o sacar conclusiones que describan las relaciones espaciales sobre los datos. Los resultados de cualquier análisis nunca serán los mismos si se realiza un reordenamiento de la distribución espacial de los elementos [130].

Se pueden describir los elementos de una Base de Datos Espacial (BDE) a través de: a) atributos: las características de los elementos que nos permiten saber que son; b) localización: permite saber donde está el objeto; c) topología: las relaciones conceptuales y espaciales entre los elementos, que mejora la interpretación del contexto y permite establecer jerarquías entre los objetos en observación [65].

Es muy importante contemplar los datos geográficos a partir de sus relaciones topológicas como concepto fundamental en un SIG. Para entender bien este término, debemos distinguir en la entidad geográfica dos tipos distintos de localización:

1. *Absoluta*: es aquella que refiere la posición de un elemento sobre la superficie de la Tierra, según un sistema de referencia explícito que son las coordenadas. Estas coordenadas son utilizadas para representar la parte geométrica de los SIG [64].
2. *Relativa*: esta noción de localización es más amplia ya que define la posición de un elemento respecto de otros elementos [64]. Encontramos aquí relaciones de tipo "junto a", "detrás de", etc. Este tipo de relación es la que permitirá interactuar los elementos entre sí para hacer análisis territoriales.

Cabe aclarar que la localización absoluta, sigue siendo relativa, ya que está relacionada con la referencia definida por convención [64], es decir se definen la latitud y longitud respecto de un punto de origen.

En un SIG los objetos tienen un componente descriptivo (nombres, datos de matrícula, fecha de creación, material que la compone, etc.) y un componente espacial (su geometría, representada por puntos, líneas y polígonos). Los puntos representan una localización en el espacio; las líneas representan un conjunto conectado de puntos; mientras que los polígonos son formados por composiciones de líneas ordenadas [67].

Se puede tener en cuenta solamente los datos alfanuméricos, sin la parte espacial. En este caso el análisis es un análisis de datos de tipo puramente estadístico. Si en cambio, nos centramos exclusivamente en la componente geométrica o espacial, se conseguirá un análisis espacial puro. Al abordar el análisis tanto con los datos alfanuméricos como con los datos espaciales, la aproximación a la problemática es de tipo territorial, estudiando la interacción de las distintas variables temáticas.

Los SIG son el resultado de adaptar de tecnologías genéricas a las necesidades específicas de datos espaciales. El proceso de estímulo y convergencia entre las nuevas tecnologías y la ciencia permitieron el desarrollo de nuevas clases de métodos analíticos y nuevas hipótesis del mundo geográfico [130].

2.2. Análisis de Datos: agrupamientos

En el agrupamiento de datos no es previamente conocida información sobre la estructura de las categorías resultantes. Se dispone de un conjunto de observaciones en el que hay que descubrir la estructura de las categorías en donde queda asignada cada observación. En esta sección vamos a dar una breve introducción al Agrupamiento de datos.

2.2.1. Definiciones y conceptos de agrupamiento

El conformar conjuntos de objetos según características similares es una técnica relevante. Conforman grupos de elementos similares, llamados también clusters o agrupamientos. Este tipo de exploración resulta de provecho para comprender relaciones complejas de carácter multivariado [66]. La localización cumple un importante rol en la determinación de un determinado fenómeno [67]. La dependencia espacial es de primer orden y de segundo orden en la mayoría de casos, lo que significa en cada zona para cada elemento son importantes sus vecinos inmediatos y los vecinos de sus vecinos [67, 69, 68].

El agrupamiento puede definirse como: dado una cantidad de n puntos ubicados en un espacio m -dimensional, se deben conformar k grupos tales que los elementos pertenecientes a cada grupo son más similares entre sí y lo más disímil a los de

otros grupos; la similitud se mide con una función distancia o alguna función de similitud [70].

El agrupamiento de datos busca cumplir con dos objetivos básicos: a) similitud intraclase: los elementos que conforman cada grupo deben ser similares entre sí dentro de ese grupo; b) distanciamiento interclase: los elementos de un grupo deben ser lo más deferente posible de los elementos de otros grupos [66]. Se debe indicar que el análisis de agrupamiento es distinto de los llamados métodos de clasificación, en donde se conoce a priori los grupos y a los cuales se asignan los elementos. El agrupamiento de datos se conforma según la similitud de los elementos, y en donde se busca conocer la cantidad y la composición de los agrupamientos [66].

Según Reyes [67] y Pascual [70], los algoritmos que agrupan datos pueden clasificarse en:

1. *Agrupación Particional*: los algoritmos agrupan los objetos según su grado de similitud, aquí encontramos a algoritmos como k-means o k-medoid.
2. *Agrupación Jerárquica*: este tipo de algoritmos va particionando el conjunto de datos por niveles, y en cada nivel se unen o dividen grupos del nivel anterior, según sea un algoritmo aglomerativo o divisivo.
3. *Agrupación basada en Densidad o Locación*: estos algoritmos tienen en cuenta la densidad de los elementos, de esta forma los agrupamientos obtenidos son aquellos que poseen una alta densidad en su interior y se separan de los otros grupos por zonas de baja densidad de elementos.

Según Soto [66] los algoritmos de agrupamiento se clasifican en:

1. *Métodos Jerárquicos*: se intenta ordenar los elementos a distintos niveles de similitud.
2. *Métodos de Partición*: en estos métodos se asigna cada elemento a cada grupo, buscando así que los grupos obtenidos sean homogéneos.

Según Pascual [70] podemos tipificar algunos algoritmos de la siguiente manera:

1. *Algoritmos Jerárquicos*: SINGLE LINK, AVERAGE LINK, COMPLETE LINK, CHAMELEON.
2. *Algoritmos Particionales*: K-MEANS, CURE aglomerativo o divisivo, PAM, CLARANS, CLARA.
3. *Algoritmos basados en Densidad*: DBSCAN, OPTICS, KNNCLUST, SNN, DENCLUE.

2.2.2. Métodos de agrupamiento

El problema de formar grupos en un conjunto de datos es muy importante en casi todo el ámbito científico. Esta conformación de grupos se realiza con los denominados métodos de agrupamiento o clustering. A continuación se describirán algunos de los métodos más conocidos.

Algoritmo K-MEANS

Es uno de los algoritmos de agrupamiento más conocidos y el más utilizado. Fue desarrollado independientemente en diferentes campos científicos por Steinhaus (1956), Lloyd (propuesto en 1957, publicado en 1982), Ball y Hall (1965) y MacQueen (1967) [125]. Su funcionamiento es simple, comienza con la definición de la cantidad de agrupamientos a obtener. Al azar se calculan k centroides (uno para cada agrupamiento), luego se asigna cada elemento del conjunto total al agrupamiento del centroide más cercano. Una vez asignados todos los elementos, se recalculan los centroides de cada grupo y se reasignan los elementos al grupo del centroide más cercano. Esto se realiza sucesivamente hasta que ya no hay cambios en los grupos de un paso a otro.

Algoritmo PAM

Este algoritmo propuesto por Kaufman y Rousseeow [127] es una evolución de K-MEANS, selecciona arbitrariamente k elementos o medoides, y va adicionando cada elemento al agrupamiento cuyo medoide sea el más cercano. Una vez realizada esta tarea, selecciona otros k elementos del conjunto total de elementos, distintos a los medoides, calcula si este nuevo elemento es más representativo del agrupamiento que el medoide, si esto no ocurre vuelve a seleccionar otro hasta que encuentra uno que sí es más representativo, entonces reemplaza el medoide por este elemento. Luego vuelve a realizar el agrupamiento en función de los medoides resultantes de la comparación. Repite estas operaciones hasta que ya no haya cambios en los agrupamientos. De esta forma conforma los agrupamientos en donde cada medoide es el elemento más representativo del conjunto. Su desventaja es su bajo rendimiento computacional.

Algoritmo CLARA

Este algoritmo es una mejora de PAM, propuesto por Kaufman y Rousseeow [128], en donde se seleccionan subconjuntos de elementos en forma aleatoria y sobre estos subgrupos se aplica PAM. Aquellos subconjuntos que tienen menor error cuadrático se seleccionan como agrupamientos. Este algoritmo es eficiente en grandes conjuntos de datos. La efectividad de CLARA depende del tamaño de la muestra; CLARA no puede encontrar un buen agrupamiento si alguno de los medoides está alejado de los mejores k -medoids.

Algoritmo CLARANS

Similar a PAM y CLARA, este algoritmo propuesto en 1994 por Ng y Han [129] selecciona aleatoriamente del conjunto un elemento E , luego selecciona para este elemento una muestra aleatoria de N vecinos (N es un parámetro dado al algoritmo). Si uno de estos vecinos tiene menor error absoluto (es más representativo del conjunto de vecinos) reemplaza el elemento E por este vecino. Repite esto hasta que no pueda conseguir un vecino mejor, es decir identificó el mínimo local. Una vez hecho esto vuelve a seleccionar en forma aleatoria otro elemento E para identificar otro mínimo local, y así sucesivamente hasta obtener los k predeterminados mínimos locales. Este algoritmo es más eficiente que PAM y CLARA.

Algoritmo DBSCAN

Es también uno de los algoritmos más usado y citado en la bibliografía, fué propuesto en 1996 por Ester et al. [126]. Para su funcionamiento se deben definir dos valores ϵ y CntMin (cantidad mínima de elementos para que una región se considere densa). El algoritmo comienza seleccionando aleatoriamente un elemento P no visitado anteriormente, si este elemento es un punto central (tiene una cantidad de vecinos mayor o igual a CntMin dentro de un radio ϵ) se comienza a construir un grupo y se agrupan en él todos los elemento alcanzables desde P (su ϵ -vecindad), es decir, sus vecinos y los vecinos de sus vecinos, esta agregación se realiza hasta completar un agrupamiento densamente conectado. Luego se visita un nuevo elemento no visitado y se repite el proceso, buscando así otro agrupamiento. Si el elemento P contiene una cantidad de vecinos menor a CntMin, el elemento es considerado ruido.

Algoritmo KAMILA

Este algoritmo es una versión escalable de k-means, adecuado para manejar conjuntos de datos de tipo mixto, propuesto por Foos et al. [71, 72] en 2016. Es un avance sobre métodos existentes en aspectos como: - las variables se utilizan con su valor de medición original, se decir no se transforman (escalan) para poder ser comparadas; - asegura un impacto equitativo tanto de variables categóricas como de las continuas [71, 72]. Combina las mejores características de dos de los algoritmos de agrupamiento más populares, el algoritmo de k-medias y los modelos de mezcla gaussiana-multinomial.

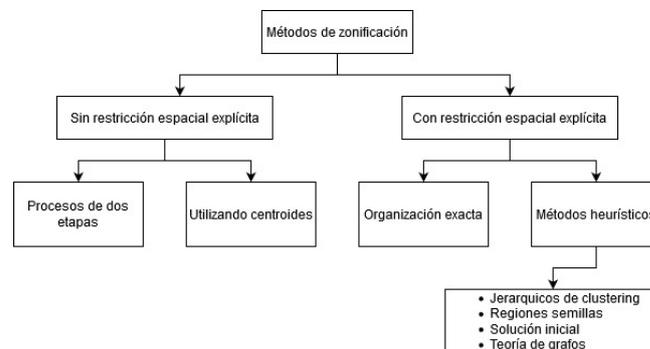


FIGURA 2.3: Tabla métodos de agrupamiento.

2.2.3. Validación de agrupamientos de datos

Existen numerosos métodos de agrupamiento y por lo general cada uno genera una solución diferente. Una de las tareas en el agrupamiento de datos es el análisis y la evaluación de los resultados obtenidos [133]. El proceso de evaluar los resultados de un algoritmo de agrupamiento es conocido como validación de agrupamiento.

Las medidas de validación pueden ser [136]:

1. *Internas*: utiliza información básica de los datos para medir la calidad del agrupamiento.
2. *De estabilidad*: evalúa la consistencia de un resultado de agrupamiento comparándolo con los grupos obtenidos después de eliminar cada columna, una a la vez.

3. *Biológicas*: evalúa la capacidad de un algoritmo de identificar agrupamientos significativos biológicamente.

Según Samundeeswari et al. [136] las medidas internas incluyen los índices Connectivity, Silhouette y Dunn, mientras que las medidas de estabilidad del agrupamiento incluyen los índices APN, AD, ADM y FOM. Las medidas biológicas incluyen los índices BHI y BSI.

Connectivity indica el grado de conexión de los agrupamientos, según lo determinado por los k vecinos más cercanos. Tanto el ancho de la silueta como el índice Dunn combinan medidas de compacidad y separación de los grupos. Las medidas de estabilidad permiten cuantificar el grado en que varían los resultados de un agrupamiento como consecuencia de eliminar, de forma iterativa, una columna o variables del conjunto de datos y comparando este resultado con el obtenido de evaluar el conjunto completo de datos. Estas últimas son relativamente costosas desde el punto de vista computacional ya que requieren repetir el agrupamiento tantas veces como columnas tenga el set de datos. Los índices de validación biológica se aplican típicamente en datos de microarrays, donde las observaciones se corresponden a genes [135].

APN (Average Proportion of Non-overlap)

La métrica APN mide la proporción media de observaciones que no son asignadas al mismo agrupamiento luego de comparar el resultado del método de agrupamiento con todas las variables disponibles contra el resultado de quitar una variable del conjunto de datos utilizadas. El valor de APN se encuentra en el intervalo $[0,1]$, siendo los valores cercanos a cero los que se corresponden con agrupamientos altamente consistentes [135].

AD (Average Distance)

La métrica AD representa la distancia promedio entre las observaciones correspondientes a un mismo agrupamiento. El valor de AD puede tener un valor $[0,\infty]$, siendo los valores más pequeños los relacionados con un agrupamiento consistente [135].

ADM (Average distance between means)

El índice ADM calcula la distancia promedio entre los centroides de los agrupamientos, utiliza en la actualidad la distancia euclidiana siendo los valores que puede tomar es índice varían entre $[0,\infty]$, considerándose los mejores aquellos más pequeños [135].

FOM (Figure of merit)

Yeung et al. implementaron este índice en 2001 [137], siendo similar al criterio de selección de modelos ESS (Error Sum of Squares). FOM es una estimación predictiva de fortaleza de un método de agrupamiento [137], calcula la varianza intra-agrupamiento.

Esta métrica adopta un enfoque predictivo y no es aplicable a todas las situaciones: si todas las condiciones experimentales contienen información independiente no es posible un enfoque predictivo [137]. Otra limitación de FOM es que no es seguro comparar los resultados de agrupamientos con diferentes números de grupos

o diferentes métricas de similitud [137]. Los valores que puede tomar FOM están en el rango $[0, \infty]$, siendo los valores más pequeños los que equivalen a un mejor rendimiento.

Connectivity

El índice Connectivity mide la distancia entre observaciones ubicadas en el mismo grupo y sus vecinos más cercanos. Este índice puede tener un valor entre cero e infinito, siendo los valores pequeños los que indican un mejor resultado [142].

Dunn

El índice Dunn [138] es una medida de evaluación de agrupamientos que permite identificar el nivel de compacidad de los agrupamientos. Un valor alto indica una mejor calidad del agrupamiento [138]. Este índice refleja la relación entre la distancia mínima entre dos grupos y la mayor distancia entre dos grupos [139]. Dunn representa la relación entre la distancia más pequeña entre las observaciones que no están en el mismo agrupamiento y la distancia más grande dentro del agrupamiento. Permite encontrar el número correcto de agrupamientos en un conjunto de datos, donde el valor máximo del índice representa la partición correcta (partición con la mayor separación entre agrupamientos y menos datos distribuidos entre agrupamientos).

Silhouette

El método Silhouette fue propuesto en 1987 por Peter J. Rousseeuw [140]. Combina los factores de cohesión y resolución. La cohesión es la similitud entre el objeto y el agrupamiento. Cuando se compara con otros grupos se llama separación. Esta comparación se logra mediante el valor de Silhouette, que se encuentra en el rango $[-1, 1]$, cuando su valor está cercano a 1 indica que existe una relación estrecha entre el objeto y el grupo. Si un grupo de datos en un modelo se genera con un valor Silhouette relativamente alto, el modelo es adecuado y aceptable [141].

2.3. Conclusiones

En este capítulo vimos el origen de los SIG y su evolución en el tiempo. Se mencionan los principales conceptos de los SIG y cuales son sus tendencias. Se habló de las herramientas desarrolladas para soportar esta teoría. Se mencionó cuales son las características que hacen a los SIG una herramienta poderosa para el análisis de datos espaciales. Se mencionan las características de los elementos que se pueden almacenar en una BDE y como considerar estas características para el análisis y obtener resultados a partir de ellas. Se describe como la ubicación espacial de los elementos observados y sus relaciones juegan un rol importante en la de un fenómeno territorial determinado.

Se expone una aproximación a la definición de agrupamiento, y se mencionan los objetivos que buscan cumplir los algoritmos de agrupamiento. Se menciona también como pueden catalogarse los algoritmos de agrupamiento de datos en: a) agrupamiento particional, b) agrupamiento jerárquico y c) agrupamiento basado en densidad. Se realiza una breve descripción de algunos de los métodos más mencionados en la literatura científica. Inmediatamente, se presenta el concepto de validación de agrupamientos de datos, indicado su importancia en el análisis de agrupamientos.

Por último, se hace una introducción a como se clasifican los índices de validación y se explican brevemente algunos de ellos.

Capítulo 3

Estado del arte

3.1. Análisis espacial de datos

En la literatura podemos encontrar numerosos trabajos referidos a los SIG en el análisis de datos. En esta sección se presenta una breve revisión de la bibliografía, centrándonos en análisis espacial estadístico y de agrupamiento. En el Cuadro 3.1 se muestra un resumen de los trabajos evaluados, tomando en cuenta el objetivo propuesto, el tipo de algoritmo empleado, y en que está basado el algoritmo.

Autor	Objetivo propuesto	Algoritmo	Metodología
Bendib [74]	Investigar como la distribución anárquica de las instalaciones públicas en la ciudad de Batna (Argelia) impactan negativamente en la movilidad motorizada	Análisis de puntos calientes (G_i^* de Getis-Ord)	Análisis estadístico
Babiker e Idris [75]	Estudiar el potencial del uso de análisis de redes para definir el área de servicio óptima de distintas establecimientos como hospitales, escuelas y la universidad de Khartoum North (Sudán)	Análisis de redes	Análisis SIG
Nesbitt et al. [76]	Estudiar la equidad distributiva de la vegetación urbana en 10 áreas urbanizadas de EE.UU.	Coefficiente de correlación de Spearman y modelos espaciales autorregresivos	Análisis estadístico
Cordes y Castro [77]	Identificar agrupamientos en los test de covid-19, en los test con resultado positivo y en la proporciones de test positivos en toda la ciudad de New York	Método I de Moran	Análisis estadístico

Adamiak et al. [78]	Estudiar la actividad de alquileres de propiedades entre pares en todo el territorio de España	Índice de Hoover	Análisis estadístico
Pérez Betancourt et al. [79]	Desarrollar una propuesta para la estratificación de territorios, basada en indicadores de salud, que facilite la incorporación del espacio en los estudios de salud pública al integrarse con una herramienta de visualización de datos espaciales.	K-MEANS	Agrupamiento
Mercadé Aloy et al. [80]	Clasificación del tejido urbano, a partir de variables derivadas de su morfología, en el Área Metropolitana de Barcelona	K-MEANS y DBSCAN	Agrupamiento
Rodríguez [81]	Estudio espacial de las situaciones de robo o hurto que sufren las personas de la localidad de Los Mártires, de la ciudad de Bogotá	K-MEANS y DBSCAN	Agrupamiento
Rava et al. [82]	Determinar las zonas de manejo homogéneas para un área de cultivo a partir de medidas de factores productivos	FUZZY MEANS C-	Agrupamiento
Yun et al. [83]	Determinar las características de humedales forestales a través del desarrollo de factores que determinen la idoneidad del sitio y la aplicación de estos factores a 107 sitios en Corea del Sur	Algoritmo de análisis jerárquico	Agrupamiento
Boampong y Çubukçu [84]	Determinar si la organización del espacio urbano en la ciudad de Izmir (Turquía) está asociado con características sociales y económicas	Algoritmo de análisis jerárquico	Agrupamiento
Basto Prieto et al. [106]	Determinar los posibles derroteros de vehículos de combate en una hipotética invasión del ejército de Venezuela en territorio colombiano	Análisis ráster, redes	Análisis SIG

Ceballos Montenegro et al. [107]	Determinar las regiones con mayor riesgo de accidente por minas antipersonales	Análisis de densidad	Análisis SIG
Matusevich [108]	determinar sitios candidatos de ubicación de centros de tratamiento y disposición de los residuos sólidos producidos por la ciudad de Corrientes	Buffers, intersecciones	Análisis SIG

CUADRO 3.1: Resumen de trabajos revisados sobre Análisis Espacial de Datos, tomando en cuenta algoritmos empleados, objetivos propuestos y metodología.

3.1.1. Análisis espacial con técnicas estadísticas

A continuación se dará un breve recorrido por algunos trabajos referidos al Análisis Espacial Basado en Estadística que se pueden encontrar en la literatura científica. Los mismos abarcan temáticas como la movilidad en una ciudad, la accesibilidad a espacios verdes en zonas urbanas, análisis de la oferta en los alquileres entre pares, análisis de agrupamientos en los resultados de los test de COVID-19.

En el trabajo de Bendib [74] se explora como la distribución anárquica de las instalaciones públicas en la ciudad de Batna (Argelia) impacta negativamente en la movilidad motorizada (movilidad, afluencia en el centro de la ciudad, congestión urbana, etc.). Para ello se integraron datos de fuentes heterogéneas, sobre los datos espaciales se realizaron tareas de corrección topológica y se especificaron en una misma proyección. Para el análisis se aplicó sobre los datos estadística espacial, más precisamente la herramienta de análisis de puntos calientes (G_i^* de Getis-Ord).

En cuanto al estudio de accesibilidad a los servicios por parte de la población podemos citar el trabajo de Nesbitt et al. [76], que con el uso del coeficiente de correlación de Spearman y modelos espaciales autorregresivos, pretende examinar la equidad distributiva de la vegetación urbana (parques) en 10 áreas urbanizadas de EE.UU.

El objetivo del trabajo de Cordes y Castro [77] es identificar agrupamientos en los test de covid-19, en los que dieron resultado positivo y en las proporciones de test positivos en toda la ciudad de New York. Se utilizó el método I de Moran para investigar la presencia de correlación en cada resultado.

Finalmente, Adamiak et al. [78] indican que los trabajos de análisis espacial realizados sobre la actividad de alquileres de propiedades entre pares por lo general se ha limitado a las grandes ciudades individuales, por ellos proponen llevar el alcance de este tipo de estudio a todo el territorio de España. Se utilizó el índice de Hoover para describir en números el nivel de concentración espacial de la oferta entre pares de vivienda, hotel y alojamiento en los municipios de España.

Estos trabajos representan algunos ejemplos de como se puede aplicar el Análisis espacial con técnicas estadísticas; y como se observará estos trabajos están relacionados con el análisis de las situaciones observadas, sobre las cuales poder delinear acciones.

3.1.2. Análisis espacial con técnicas SIG

En esta subsección se describen trabajos que se desarrollaron con metodologías propias de los SIG. Estos trabajos están aplicados en la determinación de las áreas de influencia de servicios públicos, en la determinación de vías de acceso en caso de una invasión militar, en la determinación de zonas con alto riesgo de accidentes con minas y en la determinación de zonas aptas para la disposición de desechos sólidos.

En su trabajo Babiker e Idris [75] pretenden estudiar el potencial del uso de análisis de redes para definir el área de servicio óptima de distintos establecimientos como hospitales, escuelas y la universidad de Khartoum North (Sudán). Para ello se trabajó en mejorar la topología de la caminería del área en estudio.

Basto Prieto et al. [106] intentan determinar los posibles derroteros de vehículos de combate en una hipotética invasión del ejército de Venezuela en territorio colombiano. Trabaja aquí con distintos datos para poder conformar una red que represente la existencia de caminería que permita transitar a los vehículos invasores en función de sus características técnicas, para ello utiliza un Modelo de Elevación Digital (DEM) y capas temáticas para poder determinar los posibles derroteros.

En esta línea Ceballos Montenegro et al. [107] pretenden determinar las regiones con mayor riesgo de accidente por minas antipersonales. Utilizan para ello distintos datos que incluyen DEM de área de estudio, ubicación de escuelas, centros médicos y ubicación. Sobre ellos se aplicó un análisis de densidad para determinar las zonas donde se concentran los eventos por accidentes con minas.

En cambio Matusevich [108] aplica el análisis espacial para determinar sitios candidatos de ubicación de centros de tratamiento y disposición de los residuos sólidos producidos por la ciudad de Corrientes. Este trabajo lo realizó con distintas capas que en conjunto permiten la aplicación de las restricciones que impone el problema.

Como se aprecia en estos trabajos aquí presentados, podemos darnos una idea de la diversidad de temáticas en las cuales se pueden aplicar las herramientas metodológicas exclusivas del ámbito de los SIG.

3.1.3. Análisis espacial con técnicas de agrupamiento

En esta sección se presentan algunos trabajos realizados con técnicas de agrupamiento. Los mismos se centran en distintas temáticas, como ser la seguridad, la agricultura, estudios de vegetación nativa, salud y urbanidad, si bien el espectro de aplicación en infinito.

En el caso de los trabajos que se basan en el agrupamiento de datos encontramos en Pérez Betancourt et al. [79] la utilización del método K-MEANS para desarrollar una propuesta para la estratificación de territorios, basada en indicadores de salud, que facilite la incorporación del espacio en los estudios de salud pública al integrarse con una herramienta de visualización de datos espaciales.

Mercadé Aloy et al. [80] proponen una clasificación del tejido urbano, el cual parte de variables derivadas de su morfología (calles, manzanas, parcelas, edificios), y cuyo alcance es el Área Metropolitana de Barcelona. Se aplicaron técnicas de análisis espacial para obtener los valores de las variables a utilizar y para definir las áreas homogéneas de la ciudad. Se utilizó luego el algoritmo K-MEANS, para obtener los grupos sin considerar restricciones espaciales, ya que se espera que la tipología del tejido urbano se repita en los agrupamientos que resulten.

Siguiendo la línea de los espacios urbanos, Boampong y Çubukçu [84] pretenden determinar si la organización del espacio urbano en la ciudad de Izmir (Turquía) está asociado con características sociales y económicas, para lograr este objetivo se aplicó

un procedimiento de agrupamiento jerárquico con el objetivo de agrupar vecindarios similares en cuanto a características sociales y económicas.

El trabajo de Rodríguez [81] busca estudiar espacialmente las situaciones de robo o hurto que sufren las personas de la localidad de Los Mártires, de la ciudad de Bogotá, utilizando algoritmos de agrupamiento espacial de datos, apoyándose en una infraestructura de datos espaciales. Para ello se utilizaron los algoritmos K-MEANS y DBSCAN, para determinar la ubicación óptima de las estaciones de policía y para determinar las zonas con mayor actividad delictiva respectivamente.

En el ámbito de la agricultura Rava y Santa Ana [82] pretenden determinar las zonas de manejo homogéneas para un área de cultivo a partir de medidas de factores productivos, aquí se aplica el algoritmo de agrupamiento Fuzzy c-Means, el cual es el algoritmo más utilizado de su tipo.

Respecto del medio ambiente, el trabajo de Yun et al. [83] pretende determinar las características de humedales forestales mediante el desarrollo de factores que determinen la idoneidad del sitio y la aplicación de estos factores a 107 locaciones en Corea del Sur. Para lograr este objetivo se consideraron 16 factores de cada sitio. Se analizaron las características espaciales utilizando el algoritmo de análisis jerárquico, con los que se dividió a las 107 locaciones en tres agrupamientos.

Estos trabajos presentados tienen como denominador común poder segmentar el espacio en función de las variadas temáticas en estudio.

3.2. Regionalización

En la literatura se encontró que los métodos de agregación espacial han recibido una gran atención debido a su aplicabilidad a diferentes contextos. Nos referiremos a ellos como “métodos de regionalización”, para usar un término que engloba la amplia gama de campos en los que estos métodos han sido creados o aplicados. El diseño de zonas es un problema geográfico que está presente en un amplio espectro de aplicaciones, por ejemplo la definición de distritos políticos-electorales, la definición de zonas para prestación de servicios, la zonificación para la asignación de usos del suelo, entre otros [41]. Se hace alusión a los métodos de agregación espacial de distinta manera: regionalización, diseño de zonas, zonificación, delimitación de distritos, diseño o demarcación del territorio, asignación de unidades espaciales, “clustering” espacial o geográfico, agregación espacial, partición, teselación espacial, etc [41].

La variedad de métodos de regionalización que pueden aplicarse para determinar regiones analíticas hace que sea difícil proporcionar una definición única del problema de regionalización, en la Figura 3.1, sin embargo podemos enumerar algunas características que son comunes a cualquier método que se puede utilizar para definir regiones analíticas:

- Todos los métodos unen áreas geográficas en un número predefinido de regiones mientras optimizan un criterio de agregación particular.
- Las áreas dentro de una región deben estar conectadas geográficamente (la restricción de contigüidad espacial).
- El número de regiones debe ser menor o igual al número de áreas.
- Cada área puede asignarse a una y solo una región.
- Cada región debe contener al menos un área

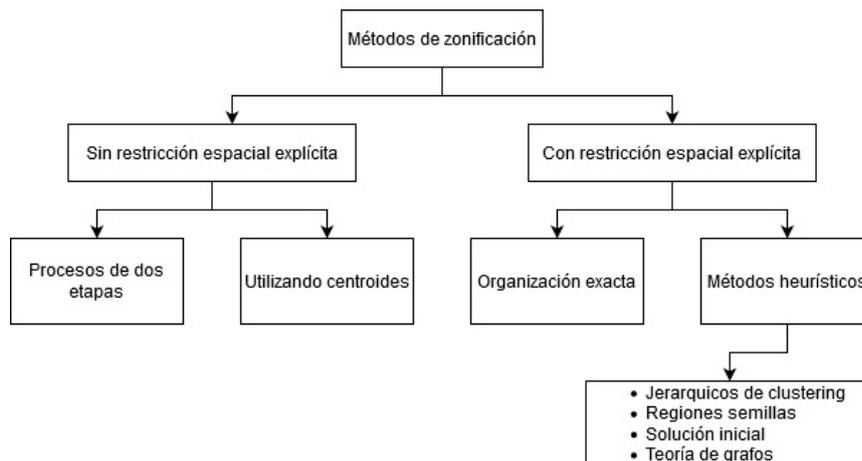


FIGURA 3.1: Taxonomía de los métodos para resolver problemas de regionalización [46].

A continuación haremos un breve recorrido en la bibliografía especializada sobre métodos de regionalización, se muestran en el cuadro 3.2 algunos trabajos que se tuvieron en cuenta.

Autor	Objetivo propuesto	Metodología
Schmidt et al. [88]	Investigar la variabilidad estacional y regional de la erosividad de las precipitaciones en Suiza	Factor R
Pérez et al. [89]	Implementación de una herramienta de software que permita la georreferenciación de pacientes con tumores malignos, la estratificación del territorio, la regionalización de servicios de salud y geosimulación	Diagramas de Voronoi Ponderados
Středa et al. [90]	Definir áreas con mayor riesgo de ocurrencia de los principales factores estresantes abióticos durante la fase crítica de crecimiento y desarrollo de los cultivos de cebada de primavera	Análisis SIG
Baxendale y Buzai [92]	Caracterizar la cuenca del río Luján a través de la regionalización de los partidos incluidos en su cuenca.	Análisis Espacial Cuantitativo
Lentes et al. [93]	Regionaliza el nivel de ingresos provenientes de la producción de leche, relacionando la duración de la estación seca con los parámetros de producción.	Análisis SIG
Ávalos Jimenez et al. [94]	Realiza una propuesta de regionalización, con base en el enfoque paisajístico y con el uso de metodologías de análisis multivariado.	Análisis SIG

Wang et al. [95]	Redefinir la línea Hu utilizando un método de regionalización automatizado GIS.	Análisis SIG (método RED-CAP)
Zolfaghari et al. [96]	Propone el uso del análisis de agrupamiento para conformar el mapa de intensidad de desertificación, y con él determinar las unidades que requieren las mismas decisiones de gestión.	Agrupamiento Jerárquico Aglomerativo
Aliaga et al. [97]	Zonificar la Región pampeana, considerando la distribución espacio-temporal de las lluvias.	Método de Ward aplicado al análisis de agrupamiento jerárquico
Braz et al. [98]	Comparar distintos algoritmos de agrupamiento para realizar una clasificación tipológica de los paisajes.	Agrupamiento jerárquico y K-MEANS

CUADRO 3.2: Resumen de trabajos de regionalización.

El trabajo de Schmidt et al. [88] tiene como objetivo investigar la variabilidad estacional y regional de la erosividad de las precipitaciones en Suiza, para ello se ampliará la red de estaciones de diaforo y se generarán mapas mensuales. Esta temporalidad de los mapas permitieron determinar regiones que se ven afectadas por la erosión de las lluvias y también aquellas que solo se ven afectadas en determinada época del año.

Siguiendo la temática de las precipitaciones, Aliaga et al. [97] proponen zonificar la Región pampeana, considerando la distribución espacio-temporal de las lluvias. En este trabajo se utiliza el método de Ward aplicado al análisis de agrupamiento jerárquico.

En la línea del uso de variables naturales para la regionalización, el trabajo de Středa et al. [90] tiene como objetivo definir áreas con mayor riesgo de ocurrencia de los principales factores estresantes abióticos durante la fase crítica de crecimiento y desarrollo de los cultivos de cebada de primavera, el mapa resultante fué elaborado en base al análisis de dos imágenes raster, combinándolas con otros datos de tipo vectorial con algoritmos SIG.

Zolfaghari et al. [96], agrega lo biológico (vegetación) a variables abióticas, proponiendo el uso del análisis de agrupamiento para conformar el mapa de intensidad de desertificación, y con él determinar las unidades que requieren las mismas decisiones de gestión. Se utiliza en este trabajo el agrupamiento jerárquico aglomerativo.

En su trabajo Baxendale y Buzai [92] utilizan variables vinculadas a conceptos sociales. En este paper se pretende caracterizar la cuenca del río Luján a través de la regionalización de los partidos de la cuenca del río, utilizando para esta caracterización el análisis espacial cuantitativo.

Lentes et al. [93] agregan conceptos económicos para regionalizar el nivel de ingresos provenientes de la producción de leche, relacionando la duración de la estación seca con los parámetros de producción. Se clasificaron los tambos según la cantidad de animales que poseen y la producción mensual de leche por animal. Estos datos se compararon con datos de temperatura y de cantidad de lluvia.

El trabajo de Ávalos Jimenez et al. [94] realiza una propuesta de regionalización, con base en el enfoque paisajístico y con el uso de metodologías de análisis multivariado. Las regiones obtenidas se basan en factores físicos, biológicos y sociales.

Partiendo de una regionalización existente, Wang et al. [95] proponen mejorarla al redefinir la línea Hu utilizando un método de regionalización automatizado GIS. Esta línea divide China en dos regiones, de dimensiones territoriales semejantes pero drásticamente distinta respecto de la densidad poblacional. Se utilizaron datos del censo de 2010 y el método REDCAP

En cuanto al desarrollo de aplicaciones informáticas, Pérez et al. [89] proponen aplicar cuatro técnicas (georreferenciación de tumores malignos, la estratificación del territorio, la regionalización de servicios y geosimulación) en una herramienta de software. La regionalización se lleva a cabo utilizando Diagramas de Vonoroi Ponderados, para incorporar en la regionalización variables que incidan en el servicio.

A diferencia de la mayoría de los trabajos mencionados, el objetivo del trabajo de Braz et al. [98] es poder comparar distintos algoritmos de agrupamiento para realizar una clasificación tipológica de los paisajes. Los algoritmos utilizados fueron de agrupamiento jerárquico y K-MEANS

Estos trabajos presentados muestran la diversidad de técnicas aplicadas para poder segmentar el territorio. Algunos de los trabajos se realizan con el objetivo de determinar problemáticas sociales o de índole climático, otras con fines económicos. Algunos trabajos se conciben con la premisa de poder segmentar por primera vez el territorio en función del objetivo propuesto. Otros parten de segmentaciones existentes con el objetivo de mejorarlas.

3.3. Conclusiones

Presentamos en este capítulo una breve revisión bibliográfica referida al análisis espacial y a la regionalización. Existen un gran número de trabajos que podemos incluir en estos conceptos, todos ellos aplicados a distintas áreas del conocimiento como la agricultura, industria lechera, inseguridad, tráfico en las ciudades, distribución de áreas verdes en ciudades, fines militares, estudios sociales y económicos, erosividad, etc.

Esto nos da una idea de la potencia del análisis espacial y de la regionalización aplicadas a: a) el análisis de situaciones relacionadas con lo territorial; b) la necesidad de segmentar el territorio en función de un objetivo; y c) la mejora de divisiones territoriales existentes.

En la primera parte de este capítulo se hizo una revisión de la bibliografía científica referida al Análisis Espacial, se clasificaron estos trabajos según las técnicas aplicadas en su desarrollo en: a) análisis espacial con técnicas estadísticas; b) análisis espacial con técnicas SIG; y c) análisis espacial con técnicas de agrupamiento. Luego se presentó una revisión de la bibliografía referida a trabajos de Regionalización.

Luego se presentó una introducción a la Regionalización, del porque de la partición del territorio por parte de las poblaciones humanas, y que consecuencias trae esto al estudiar este territorio particionado, lo que obliga a indagar en cómo encarar estudios para definir como particionar o regionalizar el espacio con el fin de lograr un mejor abordaje del territorio por parte de quienes toman decisiones.

Capítulo 4

Contexto actual

4.1. Legislación y Situación Actual

La distribución territorial de la población Argentina es heterogénea, y está definida por factores de carácter geográfico, económico, social, político, ambiental, cultural e histórico [51]. La provincia de Tucumán no está exenta de esta heterogeneidad en la distribución territorial de la población, ni los organismos de gobierno lo están de tener que organizarse territorialmente en función de sus obligaciones y necesidades. Es por ello que el ME para su normal funcionamiento administrativo-pedagógico divide a su jurisdicción según tres distintos criterios, a saber:

- En zonas definidas a partir de la Resolución N° 217/1991 del Consejo de Educación, que define los parámetros para calcular la Bonificación por Ubicación (adicional salarial de los docentes conocido como zona desfavorable).
- En zonas de supervisión, definidas a partir de la necesidad de supervisar pedagógicamente el funcionamiento de los establecimientos educativos.
- En circuitos territoriales, los cuales están definidos con la lógica de permitir a la población docente, e indirectamente a la población en general, de cada circuito desarrollarse íntegramente dentro del mismo.

Estos distintos niveles de división territorial implementados, no fueron definidos en un mismo momento, ni con criterios similares. La división territorial en función de la bonificación por ubicación, nace a partir de una normativa del ME, intenta equiparar desde lo económico las distintas realidades en las cuales está espacialmente inserta cada escuela.

La división territorial en zonas de supervisión se estableció por una necesidad natural del sistema, referida al control interno del mismo. No se encontró marco normativo que identifique las características territoriales de esta división para el nivel primario, por lo que podemos decir que en este nivel educativo la división territorial se efectivizó y se fué perfeccionando en el tiempo por uso y costumbre. Hay una división territorial de supervisión para cada nivel educativo. Se encontraron, por ser recientes en el tiempo, las normativas que implementan las zonas de supervisión para los demás niveles educativos que las utilizan.

La tercer división, en Circuitos Educativos, fue la última en implementarse. Y su aplicación obligó, en casos aislados, a redefinir algunas zonas de supervisión, para evitar el solapamiento. En otras palabras, que una misma supervisión no se intersece espacialmente con algún otro circuito. Los circuitos territoriales fueron definidos tomando como base la división política departamental de la Provincia de Tucumán, como puede observarse en las Figuras 4.1a y 4.1b. Es por ello que no se amolda a la

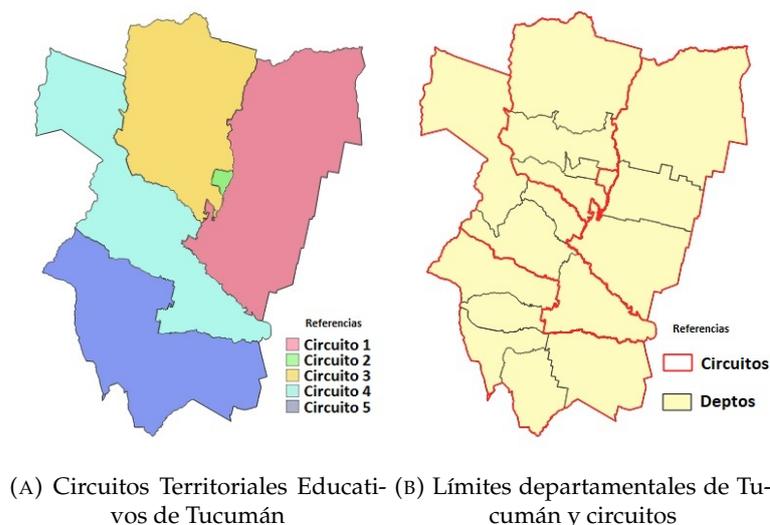


FIGURA 4.1: Comparación de los circuitos territoriales con los límites departamentales.

zonificaciones precedentes y por ello obligando (en casos muy puntuales) a realizar pequeños ajustes a lo ya existente.

El sistema educativo tucumano cuenta con: a) 485.899 alumnos (tomando alumnos de Educación Formal y de Educación no Formal) [3]; b) 1.990 establecimientos educativos (incluyendo centros unipersonales) [4]; c) 36.548 personas conforman el plantel de docentes, directivos, administrativos, auxiliares y otros [2]; d) 1.264 localizaciones escolares [4]. Estos datos corresponden a todos los niveles, tipos y modalidades de educación, tanto del sector estatal como privado, y dentro de lo estatal se encuentran los establecimientos dependientes de la provincia, de los municipios, y de las universidades.

4.1.1. Resolución N° 217/1991 (Bonificación por ubicación)

Esta resolución establece 10 (diez) parámetros a tener en cuenta para clasificar los establecimientos educativos, cada uno de dichos parámetros está constituido por variables de índole espacial, por lo tanto susceptibles de ser referenciadas geográficamente. La normativa aplica una relación entre las variables y sus valores con un puntaje determinado según el caso.

Puede indicarse que cada uno de estos parámetros corresponde a una variable o a un grupo de variables, que pueden o no ser homogéneas entre sí. Estos parámetros son los siguientes:

1. *Distancia*: otorga a cada establecimiento escolar un puntaje en función de su distancia a las ciudades de San Miguel de Tucumán (SMT) o Concepción según corresponda. Divide a la provincia de Tucumán en dos, en la zona norte las distancias se calcularán desde SMT, y en la zona sur se calcularán desde Concepción. Este parámetro es el que mayor puntaje otorga.
2. *Medio de transporte público u otros*: otorga puntaje en función de la presencia en la zona de transporte público, aquí no se tiene en cuenta la distribución espacial de los recorridos de las distintas empresas de transporte tanto urbanas como interurbanas o rurales y las áreas urbanas, sino que, por ejemplo, se

mencionan aspectos a tener en cuenta como distancia a rutas nacionales desde el establecimiento.

3. *Hospitalidad de la zona*: otorga puntaje según zonas en las cuales los caminos se tornen intransitables en época de lluvia o impliquen un riesgo en llanura y en montaña.
4. *Ubicación*: otorga puntaje en función de la distancia del establecimiento educativo respecto de un radio urbano o de una distribución de viviendas. Este puntaje puede variar para una misma distancia en caso de estar comprendida en zona de montaña o de llanura.
5. *Clima*: otorga puntaje a un establecimiento escolar según las variables temperatura máxima en verano, temperatura mínima en invierno, lluvias e inundaciones frecuentes y sequías permanentes.
6. *Salud*: otorga puntaje en función de variables referidas a la cobertura del sistema sanitario en el área de estudio, a la insalubridad permanente de la zona de influencia del establecimiento (la existencia en el territorio tanto del Mal de Chagas, como de aguas contaminadas), a la existencia o no de redes de distribución de agua potable y al suministro de energía eléctrica.
7. *Comunicaciones*: otorga puntaje según la existencia o no de distintos medios de comunicación en la zona de estudio.
8. *Seguridad*: otorga puntaje en forma binaria según existan o no servicios de seguridad. Se refiere aquí a las distintas dependencias policiales, es decir comisarías, y destacamentos policiales.
9. *Educación, cultura y recreación*: este parámetro otorga puntaje según la existencia, en el área de estudio, de instituciones educativa o de apoyo (establecimientos religiosos, ubicación de sede municipal y comunas rurales, bibliotecas, teatros o cines, clubes o centros deportivos).
10. *Economía local*: otorga puntaje según la distribución territorial de distintos factores económicos, que inciden en el bienestar de la población que se ubica en el territorio de la provincia de Tucumán.

La Res. N° 127 establece para los indicadores Hospitalidad de la zona, Clima, Salud, Comunicación y Economía local, se acumularán los puntajes de las condiciones que se cumplan. En cambio para los demás indicadores (Distancia, Medios de transporte público, Ubicación, Seguridad y Educación, cultura y recreación) no deberá cumplir con más de una condición, por lo tanto no será acumulativa en estos casos.

En la Tabla 4.1 se visualiza claramente la cantidad de puntaje máximo otorgado. Podemos observar que el mayor aportante en el puntaje total, asignado a un establecimiento educativo, es el indicador Distancia, llegando en algunos casos a representar 37,5 veces el indicador Caminos Intransitables, o 25 veces el indicador Economía Local, o casi 10 veces los indicadores Ubicación y Salud.

Debemos mencionar que esta Resolución deja mucho que desear desde el punto de vista técnico SIG, ya que plantea definiciones que llevadas a la práctica no pueden implementarse por su falta de exactitud. El ejemplo más pertinente está dado por el parámetro ocho (Seguridad) en donde se otorga puntaje al edificio educativo en función de si el área de influencia del edificio tiene o no servicios [40]. Esta definición obliga a definir en forma binaria un aspecto que puede tener una alta carga de subjetividad.

Parámetros	Puntaje máximo	Puntaje mínimo
1. Distancia	150	10
2. Medio de transporte público u otros	14	2
3. Hospitalidad de la zona	4	0
4. Ubicación	16	2
5. Clima	10	6
6. Salud	14	3
7. Comunicaciones	10	1
8. Seguridad	10	5
9. Educación, cultura y recreación	6	2
10. Economía local	6	1

CUADRO 4.1: Puntajes máximos y mínimos posibles de ser asignados.

4.1.2. Metadatos

Los metadatos geográficos son el conjunto de datos que describen la Información Geográfica y permite su óptima recuperación [61]. Son sumamente necesarios al complementar los datos que se van a compartir, ya que brindan información valiosa y necesaria para su evaluación [60]. Permiten describir el contenido, la calidad, y otras características de los datos espaciales [59]. Los metadatos son muy comunes y los utilizamos a diario por ejemplo al consultar las etiquetas de los alimentos o el prospecto de los medicamentos. Las primeras informan sobre las calorías, ingredientes, etc. que contienen, en el segundo caso las contraindicaciones, la forma de administración, etc. Actualmente en el Congreso de la República Argentina se encuentra aprobada en el Senado, y fué enviada a su tratamiento en Diputados la Ley de Etiquetado de Alimentos, que plantea la colocación de sellos frontales en los envases de alimentos con alto contenido de sodio, azúcares, grasas saturadas, grasas totales y calorías [5], el contenido de estas etiquetas son metadatos, que ayudarán a los consumidores a entender mejor los riesgos de consumir el producto en cuestión.

Los metadatos se almacenan en bases de datos distribuidas o catálogos con el objetivo de colaborar en la búsqueda de datos de índole geográfico. La visión del OGC es aquella en donde los metadatos son el centro de los servicios distribuidos de información geográfica [59]. Por ello la OGC definió los estándares de los Servicios de Catálogos, los cuales permiten el publicar y buscar en colecciones de metadatos para datos, servicios y objetos de información relacionados [62]. Los metadatos representan características de los recursos (datos) que pueden ser consultados y presentados para evaluación y procesamiento tanto por parte de humanos como de software [62].

Los metadatos pueden ser clasificados en:

- de descubrimiento: sirven para identificar el conjunto de datos de nuestro interés, permiten a quienes publican saber que datos que están disponibles. Ayudan a localizar, y entender los datos espaciales disponibles [59]
- de exploración: nos permiten identificar si los datos poseen las características que necesitamos. Nos ayudan a comprender la calidad de los datos espaciales [58]

- de explotación: permiten conocer el procedimiento de adquisición y utilización de los datos

4.2. Orígenes de los Datos a Utilizar

Los datos a utilizar tienen distintos orígenes y tipologías (puntos, polilíneas o polígonos, vectoriales o ráster). Las nuevas tecnologías de comunicación e Internet, nos permiten acceder a distintas fuentes de datos espaciales, ofrecidas en muchos casos por los propios generadores del dato, en otros casos encontramos datos que son compartidos por terceros que no generaron dicho dato. También encontramos datos generados por proyectos colaborativos como OpenStreet Map (OSM), el cual es un proyecto en donde una comunidad de usuarios mantiene en forma colaborativa una BD con datos geográficos.

Esta diversidad de fuentes, en donde a veces encontramos el mismo dato en varios repositorios, nos obliga a realizar una selección del dato a utilizar. En muchos casos no se encuentran los metadatos que nos permitirían realizar mejor esta selección.

Lo ideal sería utilizar fuentes de datos de igual calidad, como esto no es posible lo común es trabajar con las herramientas que ofrecen los SIG para mejorar los datos en su conjunto. Esta diversidad de fuentes obliga a accionar sobre ellos para procesarlos según determinados criterios con el fin de mejorar su comparabilidad entre sí, y de esta forma extraer información. A continuación se indican los datos básicos que se obtuvieron, la fuente de los mismos.

- Ubicación de establecimientos educativos, datos brindados por geoservicios del proyecto Mapa Educativo¹ del Ministerio de Educación de la Nación.
- Listado de establecimientos educativos, publicados por el Dpto. de Estadística Educativa² del ME.
- Red vial, datos obtenidos del servicio WFS de la Dirección Provincial de Catastro de la Provincia de Tucumán³ y mejorados con los datos de OpenStreet-Map⁴.
- Límites administrativos (departamentales y municipales) de la provincia de Tucumán, datos obtenidos del sitio web del Instituto Geográfico Nacional⁵.
- Circuitos educativos, confección propia a partir del mapa publicado⁶ por el Dpto. de Estadística Educativa del ME.
- Cuencas hídricas de la provincia de Tucumán, confección propia a partir del mapa publicado en el sitio web de Red de Información para el Desarrollo Productivo (RIDES)⁷.
- Modelo de elevación, datos de Earth Explorer, del Servicio Geológico de los Estados Unidos (USGS)⁸.

¹<http://mapa.educacion.gob.ar/geoservicios>

²<http://www.educaciontuc.gob.ar/sistema/basesdat.php>

³<http://www.catastro.tucuman.gov.ar/servicios/servicio-wms-web-map-service/>

⁴<https://www.openstreetmap.org>

⁵<https://www.ign.gob.ar>

⁶<http://www.educaciontuc.gob.ar/sistema/basesdat.php>

⁷<http://rides.producciontucuman.gov.ar/#descargar>

⁸<https://earthexplorer.usgs.gov/>

- Ubicación de efectores de salud de la provincia de Tucumán, datos obtenidos del sitio web del Instituto Geográfico Nacional⁹.
- Ubicación de edificios de seguridad de la provincia de Tucumán, datos obtenidos del sitio web del Instituto Geográfico Nacional¹⁰.
- Ubicación de asentamientos y edificios de la provincia de Tucumán, datos obtenidos del sitio web del Instituto Geográfico Nacional¹¹.
- Áreas programáticas del Ministerio de Salud de Tucumán, datos obtenidos del sitio de Ministerio de Salud Pública de Tucumán¹² y del sitio de Infraestructura de Datos Espaciales de Tucumán (IDET)¹³
- Unidades Regionales del Ministerio de Seguridad de Tucumán, datos obtenidos de l sitio de la Policía de Tucumán¹⁴

Para cada capa/variable se realizó un trabajo de mejora de la información de distinta índole dado que, en la mayoría de los casos, no se cuenta con los metadatos referidos a la fecha de actualización de cada una. Cada variable se trabajó en función de su tipología y las necesidades de ese dato respecto del problema a resolver. La primera acción sobre los datos recolectados fué llevarlos a un mismo sistema de referencia, se re proyectó la capa utilizando POSGAR 98 Faja 3 (posgar), según corresponde para la provincia de Tucumán.

Una vez llevados los datos al mismo SR, se procedió a trabajarlos según su tipología vectorial/raster y su necesidad dentro del problema. A continuación, se detallan algunos de los trabajos realizados sobre las capas:

4.2.1. Edificios Escolares

La capa de edificios escolares nos permite visualizar las distintas ubicaciones en donde el Ministerio de Educación imparte al menos algún nivel educativo de su oferta en cualquiera de sus modalidades. Esta capa fué confeccionada a partir del padrón de establecimientos y de la ubicación de los establecimientos educativos que se puede bajar del Ministerio de Educación de la Nación. Esta capa hace referencia a lo edilicio y no a las “escuelas”. Debemos recordar que en un mismo edificio escolar pueden convivir más de un establecimiento escolar o escuela, en diferentes turnos.

Combinando el dato de la ubicación de los edificios escolares con los datos del Listado de Establecimientos Educativos, se puede conformar la capa de escuelas (establecimientos educativos), la cual nos brindará información de la oferta educativa, su matrícula, planta de docentes, etc.

Este conjunto combinado de datos permite confeccionar distintos temáticos como ubicación de establecimientos rurales o urbanos, establecimientos privados o estatales, establecimientos de dependencia provincial, municipal o nacional, ente otros.

⁹<https://www.ign.gob.ar>

¹⁰<https://www.ign.gob.ar>

¹¹<https://www.ign.gob.ar>

¹²<https://msptucuman.gov.ar/centros-sanitarios/centros-sanitarios-establecimientos-de-salud-en-areas-programaticas/>

¹³http://idet.tucuman.gob.ar/cat_mapoteca/10-saludydesarrollosocial/#ancla

¹⁴<http://www.policia Tucuman.gov.ar/comisarias.htm>

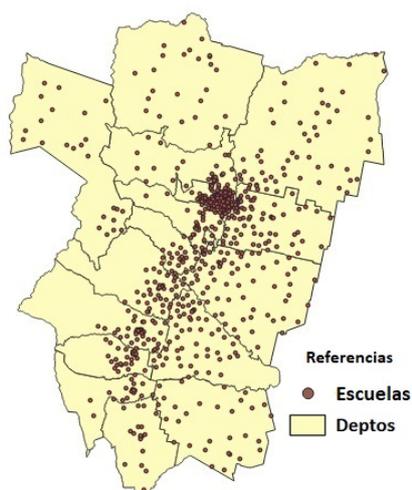


FIGURA 4.2: Localización de los edificios escolares.

4.2.2. Caminería

Las geometrías que representan la caminería (calles, rutas, caminos vecinales, etc.) de la provincia (Fig. 4.3), fueron trabajadas desde distintas fuentes. Este dato puede ser exportado completamente desde el sitio del proyecto OpenStreetMap, pero una vez realizada la descarga, se debe realizar un trabajo de limpieza de las líneas ya que conjugan en el mismo layer geometrías que se corresponden a otros elementos como ser líneas de distribución eléctrica, ríos, canales de riego, cercas perimetrales, límites administrativos, entre otros.

Como ya se contaba previamente con una capa de caminería, en el cual se habían conjugado datos del IGN, de RIDES y parte del GSMT, se optó por exportar desde OSM aquellas áreas de interés que puedan incorporarse a los datos con los que se contaba. Por ejemplo, se exportaba el área de la ciudad de Monteros, se limpiaba el dato exportado dejando sólo las geometrías correspondientes a la caminería, y se realizaba una operación de unión de Monteros con los datos de la provincia con los que se contaba previamente.

Este trabajo de ir agregando áreas de caminería de OSM a lo ya existente nos pareció más práctico que realizar una exportación total y comenzar a limpiar esta exportación. Debemos mencionar que el trabajo sobre el dato exportado no se concentra únicamente en limpiar las geometrías que no se corresponden con la caminería. En muchos casos las polilíneas que representan caminería, no se cortan en su intersección con otras polilíneas, entonces en cada intersección de las polilíneas debemos asegurarnos que las líneas/polilíneas comiencen o terminen en ese punto de intersección. Así de esta forma podemos asegurarnos que conformaremos el grafo que representa la caminería.

También se trabajó para lograr la integridad del dato, es decir eliminar geometrías repetidas y en garantizar que haya continuidad topológica en las intersecciones de las polilíneas, con el fin de conformar un grafo que permita a través de la relación de sus componentes, por ejemplo, el calcular las distancias necesarias. Se combinaron los datos de las rutas y caminerías terciarias con la red de callejeros de las distintas poblaciones de Tucumán, este trabajo es lento, ya que como se mencionó, se verifica la topología de los datos, y también hay que verificar la correcta definición

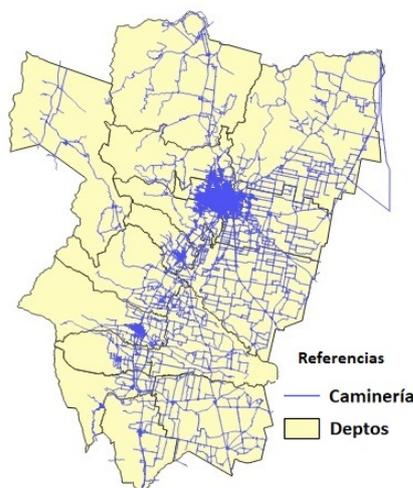


FIGURA 4.3: Mapa de la caminería de la Provincia de Tucumán.

del sentido de cada tramo de calle o ruta, lo que nos permite determinar el correcto sentido de circulación de las calles.

El dato obtenido a partir de la caminería es el que denominaremos distancia, y se corresponde con la distancia existente desde la ubicación del establecimiento educativo hasta la ciudad de SMT o Concepción, siguiendo el trazado de la caminería. Estas ciudades destino están definidas en la Resolución N° 217 [40].

4.2.3. Circuitos Educativos

En el sitio del ME¹⁵ puede observarse como está dividida la provincia de Tucumán en circuitos educativos. También se observa que cada circuito coincide en su totalidad con un departamento o un grupo de departamentos de la provincia. Entonces se replicó, partiendo de la división administrativa en departamentos de la provincia, esta capa Figura 4.1a.

4.2.4. Elevación del Terreno

Con la disponibilidad del DEM (Figura 4.4) obtenido desde el proyecto SRTM (Shuttle Radar Topography Mission), descargado del Sitio WEB de USGS, se pudo obtener tanto las curvas de nivel de la provincia, como así también se pudo calcular las pendientes del territorio de la provincia.

A partir del DEM se pudo determinar, para cada ubicación de los edificios escolares, utilizando el software QGIS, el valor correspondiente a la altura, que se corresponde en este caso con la cantidad de metros sobre el nivel del mar (msnm) a la que se encuentra el edificio escolar.

4.2.5. Cuencas Hídricas

Este dato se obtuvo a partir de la imagen publicada en el sitio de RIDES, y del servicio WMS del mismo sitio. Luego se procedió a replicar en formato vectorial tomando como base ambas fuentes de información, utilizando el software QGIS. El

¹⁵<http://www.educaciontuc.gob.ar/sistema/sistemas/mapacircuitos/>

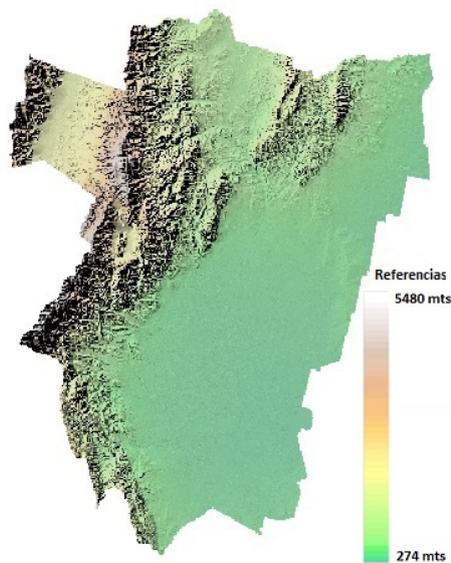


FIGURA 4.4: Relieve de la provincia

resultado puede observarse en la Figura 4.5. Así se obtuvo para cada establecimiento educativo la cuenca hídrica en donde se encuentra ubicada.



FIGURA 4.5: Mapa de las cuencas hídricas de la Provincia de Tucumán.

4.2.6. Límites departamentales

Este dato se obtuvo de filtrar para la provincia de Tucumán, la capa obtenida del sitio del IGN [63] que contiene la división político administrativa de segundo orden que incluye partido y comuna de las provincias de Argentina. Luego se transformó al sistema de coordenadas que se utiliza en este trabajo. Este dato nos permite delimitar la información correspondiente a cada uno de los departamentos de la Provincia (Figura 4.6).



FIGURA 4.6: Departamentos de la Provincia de Tucumán.

4.2.7. Límites municipales y comunales

Este dato también se obtuvo de filtrar para la provincia de Tucumán, la capa obtenida del sitio del IGN [63] que contiene la división del territorio argentino en administraciones de tercer orden que incluye comunas, juntas vecinales y demás formas de gobierno local. Luego se transformó al sistema de coordenadas que se utiliza en este trabajo, resultando el mapa que se observa en la Figura 4.7.



FIGURA 4.7: Municipios y comunas de la Provincia de Tucumán.

4.2.8. Ejidos urbanos

Se descargó la capa de ejidos urbanos del IGN, la cual representa las áreas urbanas delimitadas legalmente. Este dato se complementa con el de asentamientos

y edificios. Este dato nos permite definir a posterior características como zonas urbanas o el poder definir la existencias de determinados servicios en el territorio, en función de la envergadura de los ejidos urbanos.

4.2.9. Asentamientos y edificios

Se descargó la capa de edificación publicada por el IGN, se filtraron los datos correspondientes a la provincia de Tucumán. Este dato corresponde a la ubicación de “estructura techada relativamente permanente, diseñada para algún uso particular y que no se encuentra comprendida dentro de los demás objetos geográficos” del Catálogo de Objetos Geográficos del IGN versión 2.0 [63]. Este dato cuenta con la ubicación de más de 40.000 puntos distribuidos a lo largo y ancho de la provincia (Figura 4.8), en los mismos no se encuentran los puntos correspondientes al área correspondiente al GSMT y Tafí Viejo, como así también de los principales centros urbanos de la provincia.

Se intentó en reiteradas oportunidades bajar el dato de edificios de la Dirección General de Catastro de la Provincia, utilizando el servicio WFS brindado por la repartición, pero siempre daba error el realizar esta acción.

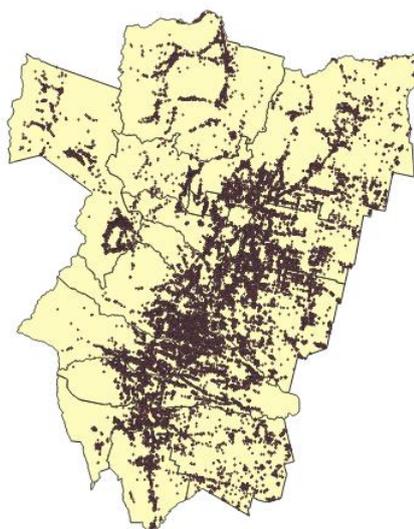


FIGURA 4.8: Asentamientos y edificios.

4.2.10. Recorrido colectivos

Se contaba previamente con un dato de elaboración propia, pero su cobertura correspondía al área del GSMT, el mismo se actualizó con los nuevos recorridos de las líneas urbanas. Luego siguiendo lo expuesto en la Ley Provincial N° 9.081, y su Decreto Reglamentario N° 897/7(STSV), se completaron los datos del GSMT con los recorridos de las líneas interurbanas de transporte de pasajeros, correspondientes al interior de la provincia.

La Resolución N° 217/1991 [40], no menciona el recorrido de las líneas de colectivos, ya que se expresa en términos de si tiene hasta la población o escuela en forma permanente o no permanente, o hasta ruta nacional. Por ello se definió un área de servicio de 1000 mts, de los distintos recorridos del transporte público de pasajeros.

4.2.11. Comisarías

Con los datos publicados por el IGN y el listado de comisarías publicados por el Ministerio de Seguridad Ciudadana (MSC) se confeccionó la capa de seguridad. Los datos del IGN no contienen la identificación de la comisaría, y los datos del MSC contienen la identificación del destacamento policial y su dirección postal, entre otros, pero no su ubicación geográfica. Se procedió a identificar la correspondencia de cada punto del IGN con cada una de las comisarías del MSC, de este cruce de datos se obtuvo la distribución territorial de comisarías.

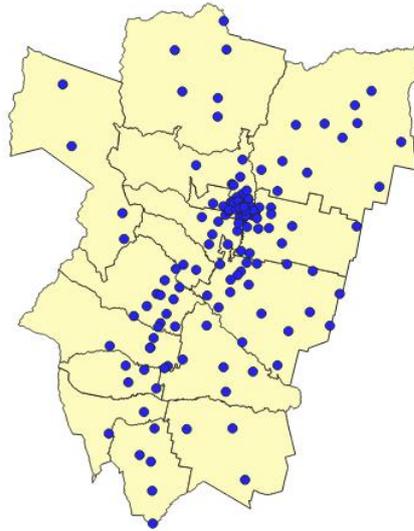


FIGURA 4.9: Comisarías.

4.2.12. Efectores de salud

Se descargó la capa de edificios de salud publicada por el IGN y se contrastó con los mapas de las Áreas Programáticas publicadas por el Ministerio de Salud de la Provincia de Tucumán (MST). En este caso se georreferenció cada archivo de formato PDF, tomando como puntos de control algunas de las ubicaciones publicadas por el IGN, previa revisión de su correcta ubicación. Una vez georreferenciado cada mapa se pudo observar una total correspondencia entre las ubicaciones del IGN con los mapas del MST. Existieron algunas inconsistencias que fueron revisadas una por una para así poder contar con el dato actualizado de los efectores de salud de la Provincia de Tucumán.

4.3. Consideraciones sobre los datos

En esta sección mencionaremos algunas consideraciones a tener en cuenta respecto de los datos. Como se mencionó, en la Resolución N° 217/1991 [40] no hay precisión en sus definiciones que permita delimitar los rangos que tendrán las variables que se emplearán. Por lo dicho se implementarán algunas definiciones con las que pretendemos salvar estas cuestiones que definiremos como vacíos legales.

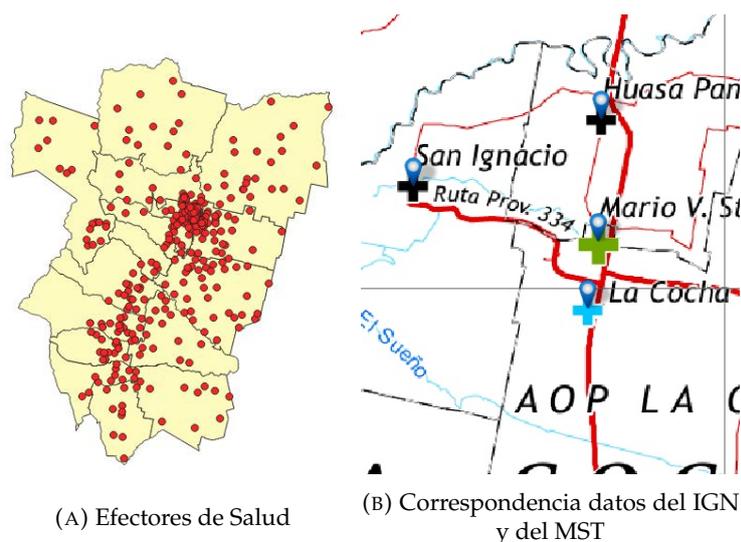


FIGURA 4.10: Efectores de salud.

4.3.1. Consideraciones generales

Como se mencionó, para la mayoría de los datos recopilados no se cuenta con la totalidad de sus respectivos metadatos, los cuales nos ayudan a comprender la calidad de los mismos [58], ya que permiten describir la información geográfica, transformándose en indispensables para evaluar y analizar los datos [100]. Hay que señalar que la exactitud de los datos espaciales depende directamente del proceso de captura utilizado para obtener dicha información [101].

Además de no contar con algunos de los metadatos, la información no está completa al 100%. Por ejemplo, si observamos un barrio conocido como El Gráfico II, al norte del Municipio de las Talitas, encontraremos datos faltantes, más precisamente la ubicación del CAM¹⁶ El Gráfico y de la Comisaria del Gráfico, tal como se puede visualizar en la Figura 4.11.

Se pueden encontrar distintos tipos de error en los datos, por ejemplo, cartografía a escalas diferentes (Figura 4.12), cartografía con diferentes sistemas clasificatorios, problemas debido a la captura en formato digital, capas de igual temática pero con diferentes sistemas clasificatorios entre otros tipos de errores [102].

Para determinar el nivel de error de los datos se debe tener en cuenta la falta de exactitud de ubicación (como ya se mencionó muy probablemente debido a un tema de escala) y el faltante de datos. El trabajo necesario para la corrección implica el revisar los datos uno a uno, o al menos hacer un muestreo representativo para determinar el nivel de error.

En algunas capas no es posible determinar el faltante de los datos obtenidos, dada el volumen de datos que contiene la misma, podemos citar la capa de Asentamientos y edificios. Si bien algunas capas pueden estar exentas de error, otras no, para este trabajo consideraremos que puede compensarse el nivel de error total aceptable.

4.3.2. Zona de alta montaña y difícil acceso

Consideramos como tal el área donde encontramos los edificios escolares ubicados en las localidades de Alto de Anfama, Ancajuli, Las Arquitas, Chasquivil, El

¹⁶Centro de Atención Municipal

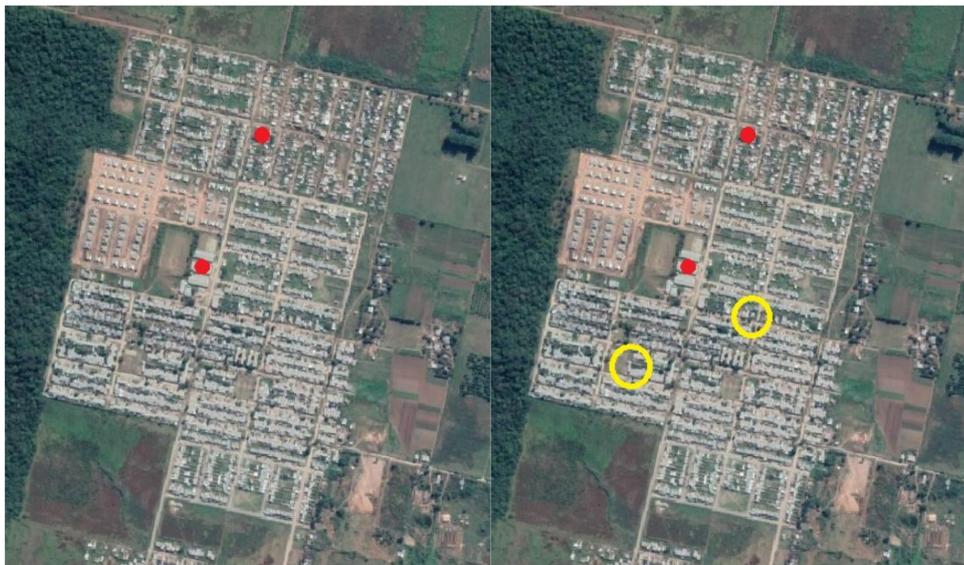


FIGURA 4.11: En rojo la ubicación de escuelas, en amarillo el faltante de ubicar (un CAM y una comisaría) en los datos del IGN.

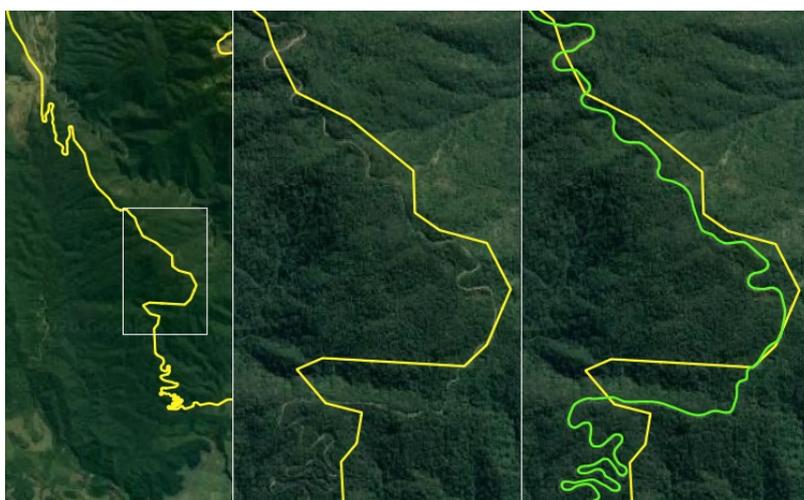


FIGURA 4.12: Error de escala.

Nogalito, Gonzalo, La Hoyada, Lara, Mala Mala, Ñorco, Rearte Sur, Rodeo Grande y San José de Chasquivil. Estas localidades se caracterizan por contar con la vía aérea como su acceso más cómodo, aunque este medio es utilizado en emergencias, se utilizan con suma frecuencia precarios caminos o sendas que pueden ser recorridos por vehículos especiales o a lomo de mulas/caballos [57]. Los caminos o sendas pueden ser transitados por vehículos dependiendo del clima, ya que en época de lluvia se vuelven intransitables.

4.3.3. Parámetro 1: Distancia

Este parámetro toma en cuenta la ubicación de los edificios escolares en el radio urbano o suburbano de SMT y/o Concepción. También considera la distancia desde el edificio escolar hasta las ciudades mencionadas (según se ubiquen hasta 20 km, hasta 50km, hasta 70km, hasta 100km y hasta 200 km).

En este parámetro nos encontramos con los conceptos de “radio urbano” y “radio suburbano”, referidos a las ciudades de San Miguel de Tucumán y Concepción. El INDEC define como población urbana a la ubicada en localidades con 2.000 o más habitantes y a la población rural a la ubicada en localidades con menos de 2.000 habitantes. Podemos suponer que en este caso hay una confusión con los términos transporte urbano y transporte suburbano de pasajeros, que hacen referencia a los recorridos del transporte, en el primer caso se desplazan dentro de los límites de la ciudad, en el segundo caso se refieren a aquellos recorridos que salen de los límites de la ciudad, pero estos términos consideran a la ciudad como una sola unidad tomándola como referencia de los recorridos del transporte público de pasajeros.

Entonces, a partir de lo conceptualizado por el INDEC¹⁷ y de la terminología referida a si los recorridos se realizan dentro del radio urbano, o si se extienden más allá del radio urbano, consideramos que las ciudades de San Miguel y Concepción, cuyas poblaciones superan los 2.000 habitantes, se corresponden íntegramente con el concepto de radio urbano que es indicado por la Resolución N° 217.

4.3.4. Parámetro 2: Medios de Transporte Público u otros

Este parámetro toma en cuenta si la población o escuela cuentan con transporte público, en forma permanente o no permanente. Considera también el recorrido de transporte público hasta rutas nacionales y si el edificio se encuentra dentro de determinadas áreas de servicio de la ruta nacional. Por ello se confeccionó un buffer de 1.000 mts sobre los recorridos de los colectivos, considerando que si el edificio escolar se encuentra dentro del mismo cuenta con el servicio de transporte si es que este edificio escolar no se encuentra dentro de algún ejido urbano, ya que consideramos que no existe ejido urbano en la provincia sin el servicio de transporte público de pasajeros.

La Ley N° 6.210 regula el servicio del transporte público de pasajeros en la Provincia de Tucumán. En la misma se establece que el transporte de pasajeros estará sujeto a concesión cuando reúna las características de oneroso, público, regular y periódico. El Decreto N° 2129/91, que reglamenta la mencionada ley, define como regular aquellos servicios con recorridos, frecuencias y horarios constantes y previamente establecidos. La ley N° 6.210 también define los “servicios especiales”, los cuales incluyen, entre otros, aquellos que satisfacen necesidades derivadas de las relaciones entre instituciones de distinta índole con su personal. Consideraremos aquí a aquellos medios utilizados para el transporte hasta el lugar de trabajo (en este caso la escuela). No encontramos que se aplique el concepto de “servicio especial”, tal como hemos considerado, en el ámbito educativo. Por lo anterior, las características que establecen la Ley N° 6.210 y el Decreto N° 2129/91 nos permiten definir que no se cuenta con medios de transporte públicos en forma no permanente en el ámbito de la Provincia de Tucumán.

4.3.5. Parámetro 3: Hospitalidad de la Zona

Este parámetro toma en cuenta si el edificio escolar se encuentra ubicado en zonas con caminos intransitables en época de lluvia o con caminos de riesgo, tanto en llanura como en montaña.

No se encontró información de fuente confiable para poder aplicar este parámetro

¹⁷<https://www.indec.gob.ar>

4.3.6. Parámetro 4: Ubicación

Este parámetro toma en cuenta si el edificio escolar se encuentra hasta cinco kms del radio urbano, o si existe una distribución de viviendas en el área de servicio del establecimiento educativo (con rangos de hasta tres kms, hasta cinco kms y seis o más km), considerando esto último para zona de llanura como de montaña.

En este trabajo se considerará como "distribución de viviendas.^a un conjunto de 5 o más asentamientos o edificios. Es decir que cada asentamiento o edificio será considerado como una vivienda.

Respecto de la división del territorio en llanura o montaña, se realizará respecto de la curva de nivel que representa los 950 msnm, ya que es en torno a esta que se produce un cambio de la pendiente haciéndose más pronunciada por arriba de ese valor. Siendo considerados ubicados en zona de montaña a aquellos edificios escolares que se encuentran a 950 o más msnm. Aquellos que se encuentren por debajo de 950 msnm, serán considerados ubicados en zona de llanura (Figuras 4.13a, 4.13b y 4.13c).

Se tendrá como zona aislada o semiaislada la zona correspondiente a los establecimientos de alta montaña y de difícil acceso.

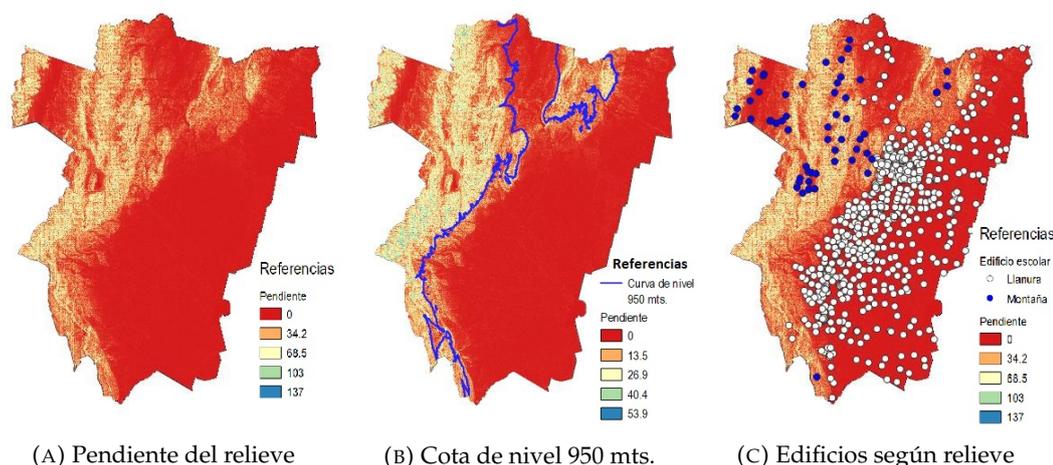


FIGURA 4.13: Datos obtenidos a partir del DEM.

4.3.7. Parámetro 5: Clima

Este parámetro toma en cuenta si el edificio escolar se encuentra ubicado en una zona con temperaturas menores a 5° en invierno, temperaturas mayores a 35° en verano, si se ubica en zonas inundables o con sequías permanentes.

Se aplican las variables que se encontraron en los repositorios especializados, se trabajó con valores de temperaturas medias para enero (enero) y julio (invierno), por no encontrarse datos para valores medios máximos y mínimos en los periodos correspondientes. Respecto de la variable sequía se tomó la definición que encontramos en el sitio del INTA¹⁸ donde se explica la diferencia entre sequía y aridez, "Debe comprenderse que la aridez (fenómeno estructural climático), difiere de la sequía (fenómeno coyuntural); mientras esta última se manifiesta en el tiempo (períodos secos), la aridez constituye un fenómeno espacial representado en las regiones áridas. El término desierto, se reserva para identificar un paisaje denudado a causa de

¹⁸ // inta.gob.ar/noticias/desertificacion-en-argentina-el-problema-de-las-60-millones-de-hectareas

la sequía y la aridez". Por esto la desertificación es una consecuencia de la sequía y de la aridez, entonces las zonas con sequías permanentes podemos determinarlas con la capa de riesgo grave y severo de desertificación que se publica en el sitio de RIDES¹⁹

4.3.8. Parámetro 6: Salud

Este parámetro toma en cuenta si el edificio escolar se encuentra en una zona con cobertura del servicio de algún efector de salud. También considera si el establecimiento se encuentra en una zona que cuenta con el servicio de la red eléctrica y/o la red de agua potable.

Se considera que un establecimiento educativo tiene servicio (hospitalario, dispensario o posta sanitaria) si el mismo se encuentra ubicado dentro del área de responsabilidad de un efector sanitario cuya tipología corresponde a las mencionadas en la Resolución N° 217. Existen en la Provincia de Tucumán distintos tipos de efectores sanitarios, como los hospitales, CAPS, CAM, CIC, CAC y postas sanitarias. El término dispensario no es tenido en cuenta actualmente dentro del Sistema de Salud, por lo que se realizó una equivalencia entre la tipología de efectores del sistema de salud de Tucumán, y la tipología de efectores mencionados en la Resolución N° 217.

- Se consideró como dispensario a los CAPS y CAM - Se consideró como posta sanitaria a las postas sanitarias, a los CIC y a los CAC

Para la determinación de las áreas con o sin el servicio de agua por red de distribución, se consideró que los ejidos urbanos en su totalidad cuentan con este servicio, considerando también que el resto de la provincia no cuenta con un sistema de distribución de agua, dada la dispersión de las viviendas.

Se considera que no tiene luz eléctrica cuando el establecimiento se encuentra en una zona fuera de la red de distribución de energía.

4.3.9. Parámetro 7: Comunicaciones

Este parámetro toma en cuenta si el edificio escolar se encuentra en una zona que cuente con algún tipo de medio de comunicación.

Se consideró en este caso que las áreas urbanas poseen todos los medios de comunicación incluido el servicio de Internet que está brindado por la telefonía móvil. Se considera que los establecimientos de alta montaña y de difícil acceso poseen solamente el servicio de radio comunicación. Las zonas no incluidas en las anteriores se consideran con el servicio de telefonía. Se considera que los servicios de cabinas telefónicas y telégrafo, correo y/o estafeta se encuentran en los centros que poseen todos los servicios. Este punto se estima se definió para los tiempos en que la falta de algunos de los servicios mencionados hacían más lentas las comunicaciones, hoy en día las nuevas tecnologías se impusieron, y el faltante de alguno de estos servicios no hace mella en la velocidad de la comunicación.

4.3.10. Parámetro 8: Seguridad

Este parámetro toma en cuenta si el edificio escolar se encuentra en una zona que tiene seguridad o no.

¹⁹<http://www.rides.producciontucuman.gov.ar/>

Se consideró en este caso la conceptualización del parámetro 6, es decir si el edificio escolar se encuentra o no dentro del área de responsabilidad de comisaría alguna. Como todo el territorio de la provincia está dividido en áreas de responsabilidad de las distintas comisarías o dependencias policiales, se considera asignar para este parámetro a cada edificio escolar el valor correspondiente a "Tiene servicios".

4.3.11. Parámetro 9: Educación, Cultura y Recreación

Este parámetro toma en cuenta si el edificio escolar se encuentra en una zona que tiene instituciones de apoyo (iglesia, municipalidad o comuna rural, biblioteca, teatro o cine, clubes, complejos o centros deportivos).

Se consideraron las zonas de alta montaña como que no poseen instituciones de apoyo (las indicadas en la normativa), Las ciudades cabeceras de departamento, las zonas veraniegas y las grandes áreas urbanas se consideran como que poseen todas las instituciones de apoyo. Quedando el resto de la provincia como zonas con algunas instituciones de apoyo.

4.3.12. Parámetro 10: Economía Local

Este parámetro toma en cuenta el tipo de economía predominante en la zona donde se encuentra el edificio escolar. Considera que la economía local esté basada en la industria o comercio, en la agricultura, en la ganadería o sea una economía familiar o de subsistencia.

Se consideraron las zonas de alta montaña como zonas con economía familiar, ya que las mismas se encuentran alejadas de los centros urbanos, sin buenas vías de comunicación por donde hacer llegar su producción hasta los potenciales consumidores. Se considera las zonas urbanas como centros que poseen economías que dependen de las industrias, comercio, servicios y turismo. Las zonas que no entran en las categorías anteriores se consideran dependientes de economías de agricultura o ganadería (incluyendo en este caso la láctea, es decir los tambos)

4.4. Acondicionamiento de los datos obtenidos

Los distintos datos que se obtuvieron nos permiten generar otros datos, en función de las necesidades de este trabajo. Para ello haremos uso del potencial de los geoprocetos que se pueden aplicar a los datos geográficos. Muchos de los datos mencionados en los parámetros de la Resolución N^o 217, pueden ser obtenidos a partir de los datos que ya se obtuvieron.

4.4.1. Distancia

Este dato se obtuvo a partir de la capa de caminería. Se utilizó el paquete pgRouting que es una extensión que añade funcionalidad de enrutamiento y de análisis de datos a una base de datos PostGIS/PostgreSQL. La funcionalidad utilizada fué el procedimiento de cálculo de distancia que aplica el algoritmo Dijkstra. Para esto fué necesario que la caminería conforme un grafo sin nodos ni arcos repetidos (trabajo descripto anteriormente), y que en cada nodo haya continuidad topológica.

4.4.2. Distancias a rutas nacionales

Este dato se obtuvo también a partir de la capa de caminería. Se filtraron la geometrías correspondientes a las rutas nacionales, y luego se calcularon buffers o áreas de servicio de hasta 1 km de la ruta, de entre un km y diez km, de entre diez km y 20km y entre 20km y 30km.

4.4.3. Distancias desde la escuela

Este dato se cálculo a partir de la capa de edificios escolares. Para cada punto que representa la ubicación de los distintos edificios escolares se generaron distintos buffers, uno de entre uno km y tres kms, otro de entre tres y seis kms. Este dato nos permitirá calcular el puntaje correspondiente al parámetro 4 de Ubicación de la Resolución N^o 217 [40]

4.4.4. Cobertura del servicio de Red Eléctrica

Este dato se conformó a partir de la capa de asentamientos y edificios. Se calculó para cada punto de esta capa un área de cobertura de 150 mts. Esto nos definió zonas donde estas áreas de cobertura se superponían, se consideraron que las manchas de estas áreas que se extendían siguiendo rutas o caminos corresponden a zonas con cobertura de la red eléctrica. También se hizo esta consideración para las manchas sobre zonas urbanas y zonas cercanas a las mismas. Se descartó la zona de alta montaña y difícil acceso de estas consideraciones. Así de esta forma se fué conformando el área servida por la red de energía eléctrica.

4.4.5. Montaña y llanura

A partir del DEM se pudo dividir a la provincia de Tucumán según su relieve en dos regiones, la región de llanura y la región de sierras y cerros. Esta partición del territorio se realizó tomando en cuenta el área territorial donde se evidencia un cambio abrupto de pendiente, es decir de una pendiente suave o mínima a una pendiente pronunciada. Esta situación se observó que se presenta mayormente en torno a la curva de nivel de 950 mts.

El cálculo de pendiente se realizó con el software QGIS, utilizando la librería GDAL que está incorporada en QGIS. De esta forma se pudo definir las áreas que definiremos, según lo establecido en la Resolución N^o 217, como zona de montaña y zona de llanura.

4.4.6. Almacenamiento de los datos

Los datos descritos, en su totalidad fueron almacenados en una BD geográfica, se utilizó el motor de BD PostgreSQL, a la cual se añadió PostGIS. Este complemento amplía las funcionalidades de PostgreSQL, permitiendo almacenar, consultar y procesar datos de tipo geográfico. QGIS nos permite trabajar con datos almacenados en una BD PostGIS.

La capacidad de poder almacenar en una BD permite acceder y operar los datos espaciales desde desarrollos propios, ampliando de esta manera el potencial que nos brinda la herramienta QGIS. Por ejemplo, podemos acceder a los datos a través de scripts para calcular la distancia entre dos puntos (utilizando PGRouting por ejemplo).

4.5. Conclusiones

En este capítulo hicimos un acercamiento a la normativa que utiliza el ME de Tucumán, y que es la base de este trabajo. Del análisis de la misma podemos identificar las necesidades de datos que utilizaremos, también poder definir que datos son producto de aplicar procesos sobre los datos obtenidos. Una vez definidos los datos que utilizaremos, se procedió a la búsqueda de los mismos, investigando en distintas fuentes de datos libres que existen y que son pertinentes de utilizar. Se hace una introducción al concepto de metadato y su importancia para la búsqueda y para analizar las características de los datos encontrados. Se indica para cada caso las fuentes de los datos y su tipología, también el proceso mínimo necesario para acondicionar cada dato. A posterior de contar con los datos espaciales que pueden accederse, se definieron algunas consideraciones a tener en cuenta para poder cumplir con las definiciones establecidas en la normativa. Se indicaron los tipos de errores que encontramos en los datos, y como se solucionaron algunos de ellos. Se estableció la definición de Zona de alta montaña y difícil acceso, que esta presente en la normativa y es necesario de aplicar. Se indica como se acondicionaron los datos y como se procesaron algunos de ellos para obtener otros datos que pueden derivarse de los primeros.

Capítulo 5

Resultados obtenidos

En este capítulo se presentará el tanto los resultados obtenidos de los procesos llevados a cabo sobre los datos con el fin de cumplir con el objetivo del presente trabajo. Comenzaremos con el cálculo del puntaje que otorga la Resolución N° 127/1991 [40], esta variable será incorporada al dataset a utilizar para obtener la redefinición de los Circuitos Educativos, esto se esquematiza en la Figura 5.1.

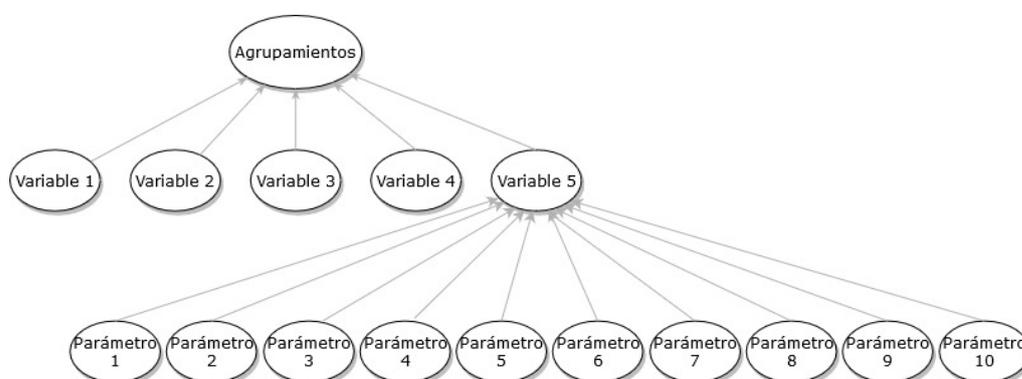


FIGURA 5.1: Vinculación entre las distintas variables utilizadas.

5.1. Determinación del Puntaje para el Cálculo de la Bonificación por Ubicación

Tomando como base el trabajo de Olea y Araujo [117], se programó un script en PHP para calcular el puntaje total que corresponda a cada edificio escolar, siguiendo los lineamientos definidos en la Res. N° 217 [40].

La resolución mencionada define 10 parámetros a tener en cuenta, los cuales pueden definirse a su vez como capas SIG, algunas con características distintas para la misma variable, o compuesto de variables distintas, con distintas características para cada una de ellas. Por esto se puede indicar que se están normalizando y ponderando las variables, según el criterio tomado en cuenta al momento de la redacción de la resolución.

El script que se desarrollo hace uso de procesos geospaciales o geoprosos, que consisten en una serie de análisis basados en el procesamiento de la información geográfica, están destinados a establecer relaciones entre dos o mas capas independientemente de su naturaleza. Estos geoprosos pueden realizar operaciones como extraer o superponer datos, reproyectar capas, establecer rutas óptimas, etc.

Una instrucción SQL que realiza una consulta en una BD es muy ágil y poderosa respecto del volumen y la magnitudes que los datos que tratará, también al momento de manejar datos espaciales es mucho más ágil que un SIG de escritorio

en términos de cantidad de tiempo requerido para realizar un proceso en un volumen dado de datos que pueden ser consultados/procesados. Por ello todas las consultas se resolvieron con PostGIS, esta extensión de PostgreSQL contiene cientos de funciones disponibles, algunas pueden generar nuevas geometrías a partir de otras (buffers, unir las, interseccionarlas, etc.), otras permiten consultar atributos de las geometrías (longitud, superficie, perímetro, etc.).

Los cálculos necesarios se realizan directamente en el motor de BD seleccionado, logrando que los tiempos de ejecución de análisis SIG basados en PostGIS sean notablemente más cortos y eficientes que cualquier otro SIG de escritorio. Cada consulta calculará y/o devolverá el puntaje que se asigna a cada coordenada según lo especificado en la normativa.

Ejemplo

Se calculará el puntaje para un edificio escolar ubicado en las coordenadas (-65,6473 ; -27,2632) las que se programaron en el script mencionado. La lógica de cálculo se ejemplifica en la Figura 5.2.

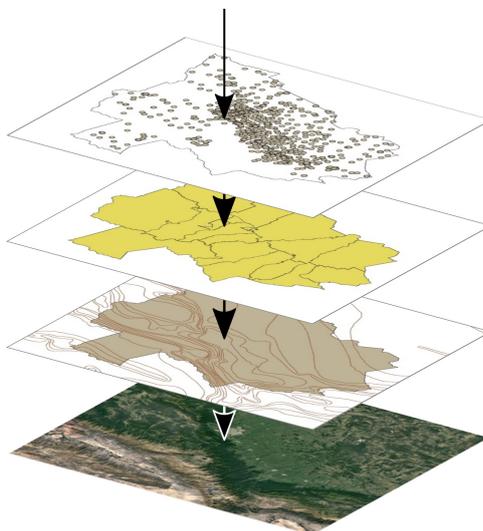


FIGURA 5.2: Se calcularán los parámetros correspondientes a la coordenada indicada.

El cálculo se hace por cada variable que indican los distintos parámetros, a través de consultas SQL espaciales. Utilizando la potencia de combinar PostgreSQL con PostGis, se puede hacer una consulta con el valor XY de la coordenada, pudiendo en la misma verificar el resultado de interseccionar la mencionada coordenada contra una figura geométrica de tipo polígono, obteniendo si esta intersección es verdadera o falsa. Para el primer caso (verdadero), se comprueba que el punto corresponde al interior del polígono, y por lo tanto le corresponden las propiedades del mismo.

Siguiendo el ejemplo anterior, para los distintos grupos de parámetros, se encontrarán distintos resultados de estas comparaciones, pudiéndose asignar un puntaje para cada caso, los cuales se muestran a continuación:

- *Distancia*, esta debe ser calculada a la ciudad de Concepción, por ubicarse la coordenada en el departamento Chicligasta. Como la distancia más corta, siguiendo la red caminera, es de 16,84 Km, se corresponde con el ítem "Escuela ubicada hasta 20 Km de San Miguel de Tucumán o Concepción", recibiendo por ello una cantidad de 30 (treinta) pts.

- *Medios de transporte público u otros*, para este parámetro se determina que la localización en observación cumple con encontrarse a una distancia de hasta 10 Km de una ruta nacional o apeadero, con lo que corresponde asignar por este parámetro una cantidad de 8 (ocho) pts.
- *Hospitalidad de la zona*, como se indicó, para esta definición no pudimos establecer las áreas especificadas en la normativa, por tal motivo a todos los edificios se les asignará un puntaje de cero puntos.
- *Ubicación*, no se encuentra a una distancia menor de cinco kms de un radio urbano, por lo tanto se procede a analizar las siguientes reglas de este parámetro, con lo que se identifica que la localización en estudio se encuentra en zona de llanura. Luego se analiza su cercanía a alguna distribución de viviendas, como se determina que existe un conjunto de viviendas en un radio de entre uno y tres kms, el puntaje que corresponde para este parámetro es de 6 (seis) pts.
- *Clima*, para este parámetro la ubicación en estudio permite identificar que no se ubica en la franja de temperatura inferior a los 5º (cinco grados) de temperatura para invierno, por lo que, según la normativa le corresponde un puntaje de cero puntos.
- *Salud*, las coordenadas que representan la ubicación en estudio se encuentran dentro del área de responsabilidad de un efector, el CAPS Sargento Moya, como el mismo se corresponde con un CAPS por lo que le corresponden 3 pts, y además esta ubicación se encuentra en un área donde encontramos que existe un servicio de distribución de energía eléctrica en baja tensión, otorgando por ello 1 pto. Por lo antes expresado, para este parámetro, le corresponde un puntaje de cuatro puntos.
- *Comunicaciones*, el análisis de este parámetro, indica que no se encuentran ni en zona de alta montaña y difícil acceso, ni en zona urbana o zona turística, por lo tanto corresponde asignar un punto por ubicarse en una zona con teléfono.
- *Seguridad*, la ubicación en estudio contrastada contra el área de influencia del servicio de seguridad, nos permite identificar q se encuentra con cobertura de seguridad, por lo tanto le corresponde un puntaje de 5 (cinco) pts.
- *Educación, cultura y recreación*, la ubicación cumple con tener al menos una institución de apoyo la zona, corresponde asignar por este parámetro un puntaje de cuatro puntos
- *Economía*, la ubicación en estudio se encuentra en el área definida como de economía basada en agricultura o ganadería, ésta situación permite asignar a la misma un puntaje de seis puntos.

El Cuadro 5.1 indica el puntaje que el script devuelve para la localización indicada, según los distintos parámetros analizados.

Debemos recordar que este script, si bien aquí está ejemplificado para un edificio escolar, se programó para calcular el puntaje que le corresponderá a todos los edificios escolares.

Concepto	Puntaje asignado
1) Distancia hasta Concepción es de 16,085 kms	30
2) Medio de transporte público u otros	8
3) Hospitalidad de la zona	0
4) Ubicación	6
5) Clima	0
6) Salud	4
7) Comunicaciones	1
8) Seguridad	5
9) Educación, cultura y recreación	4
10) Economía local	2
Puntaje Total	60

CUADRO 5.1: Ejemplo de resultado obtenido para una coordenada en particular.

5.2. Métodos basados en densidad

5.2.1. DBSCAN

El primero en ser utilizado fue el algoritmo basado en densidad denominado DBSCAN. Este propone que para incorporar un punto en un agrupamiento, tienen que existir un mínimo de observaciones dentro de un radio de proximidad pequeño, entonces cada agrupamiento obtenido estará separado por espacios vacíos o pocas observaciones. Este método encuentra los agrupamientos que se conforman según los parámetros con que se corre el algoritmo. El inconveniente de utilizar este método es que no se puede indicar la cantidad de agrupamientos deseados, ya que el algoritmo mismo los detecta, otro inconveniente de este algoritmo fué poder determinar el valor de ϵ , necesario para el funcionamiento de este algoritmo.

Las características de la dispersión del universo de observaciones (escuelas) que encontramos en la provincia de Tucumán, hace que este algoritmo tienda a agrupar los puntos de la zona pedemontana en un único cluster. La ciudad de SMT, por la

densidad de locaciones que posee, sería parte de este cluster. Por esta razón, previamente eliminamos estos puntos previendo éste comportamiento del método.

En las Figuras 5.3a, 5.3b, 5.3c, 5.3d, 5.3e y 5.3f se pueden observar los resultados de algunas ejecuciones del algoritmo, siempre obteniendo 5 (cinco) agrupamientos. Los puntos azules corresponden a los puntos ruido, que no son considerados en los agrupamientos del resultado.

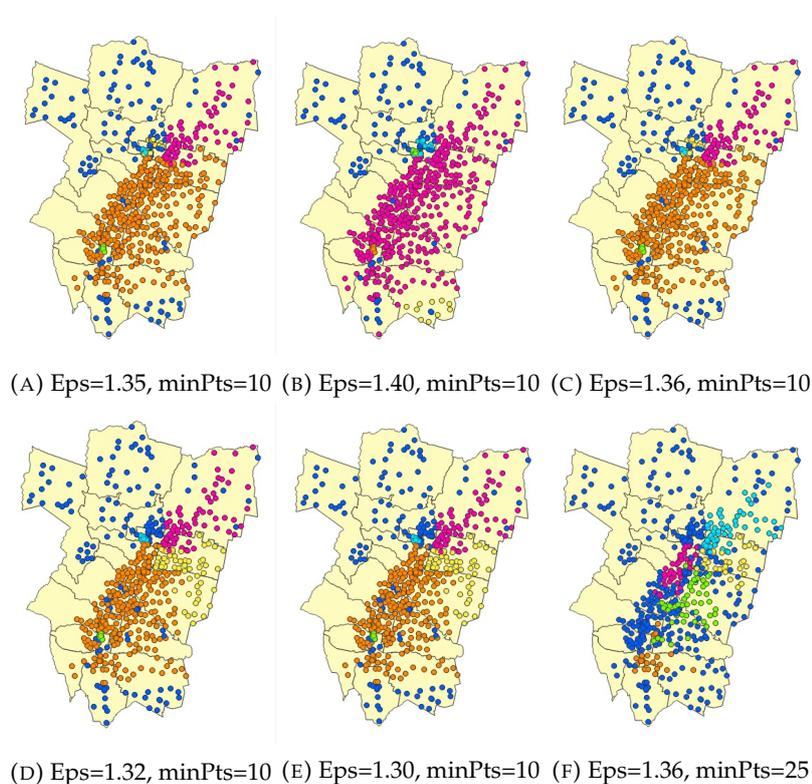


FIGURA 5.3: Resultados del método DBSCAN con la totalidad de datos.

5.3. Métodos de particionado

5.3.1. K-MEANS

De todos los algoritmos presentados hasta aquí, K-Means es posiblemente el más popular entre los algoritmos de clústeres. Es un algoritmo simple y rápido, y tiene la capacidad de tratar eficientemente datasets muy grandes. Este método de agrupamiento particiona un conjunto de elementos en k grupos, en donde cada observación (en este caso los edificios escolares) pertenece al grupo cuyo valor medio es más cercano. K-Means es un algoritmo sensible al ruido y los outliers debido a que un número pequeño de los datos pueden sustancialmente influenciar el valor medio (centroides de cada clúster).

Se trabajó con este algoritmo para obtener 5, 6, 7, 8, 9 y 10 clusters. Las variables utilizadas fueron puntaje total, ámbito, departamento, municipio, cuenca hídrica, elevación, distancia, latitud, longitud. Los resultados obtenidos no fueron satisfactorios, ya que si bien se observa una agrupación espacial, muchos puntos de distintos clusters se encuentran dispersos por el territorio (Figuras 5.4a, 5.4b, 5.4c, 5.4d, 5.4e y 5.4f).

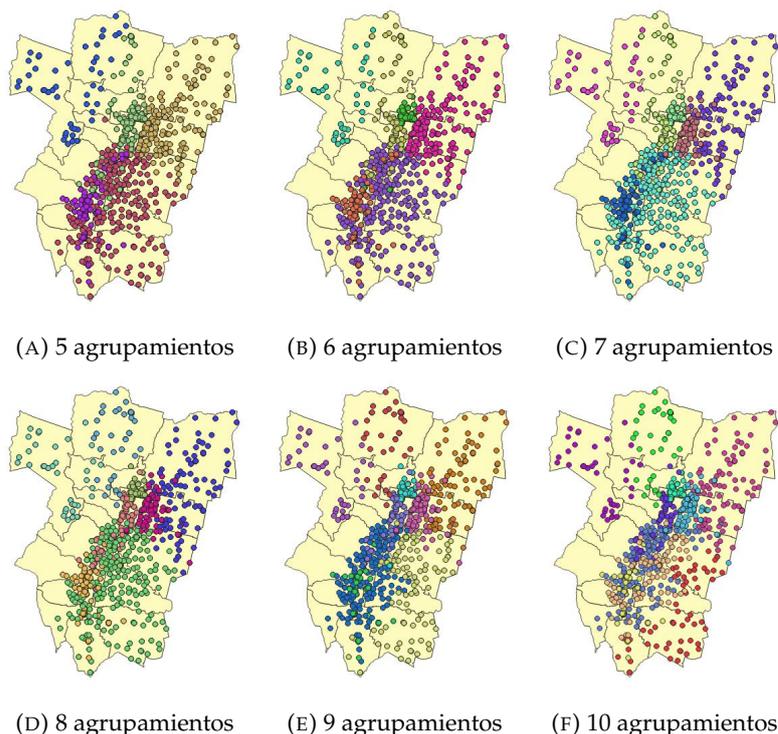


FIGURA 5.4: Resultados del método k-means con la totalidad de datos.

Se realizaron a posteriori diversas corridas con el objeto de identificar variables que incidan en esta dispersión de los clusters en el plano. Se fué probando al descartar las variables municipio, cuenca y ámbito, ejecutando nuevamente el algoritmo con el resto de las variables. Al observar los resultados en cada caso se pudo observar la tendencia de un sesgo por alguna variable q a posteriori quitaríamos del set de datos. Como ejemplo mostraremos en las Figuras 5.5a, 5.5b, 5.5c, 5.5d, 5.5e y 5.5f. Algunos resultados obtenidos donde puede observarse un sesgo que se puede vincular con la variable departamento, por ejemplo el departamento Chicligasta -ubicado en el sudoeste de la provincia de Tucumán- figura vinculado o incluido en su totalidad en clusters que no son contiguos o figura como un agrupamiento independiente, y también en el agrupamiento que se encuentra al noroeste de SMT, y que comprende las ciudades de Yerba Buena, Tafí Viejo, Las Talitas, Los Pocitos y localidades que se encuentran entre ellas. En los algoritmos anteriores se conformaba un único agrupamiento rodeando la ciudad de SMT, en este caso no, y se separan estos agrupamientos siguiendo el límite administrativo departamental.

Por último, se ejecutó el algoritmo solamente para las variables puntaje total, elevación, distancia, latitud y longitud, dejando de lado las variables ámbito, departamento, municipio, cuenca y zona. El resultado obtenido puede indicarse como aceptable dado que los agrupamientos son bastantes consistentes y bien definidos entre sí (Figuras 5.6a, 5.6b, 5.6c, 5.6d, 5.6e y 5.6f).

Observando los resultados del algoritmo K-Means con las variables puntaje total, elevación, distancia, latitud y longitud, advertimos un patrón en los resultados. Se generan cuatro agrupamientos que se mantienen independientemente de la cantidad de clusters que se obtienen. El primero se encuentra en el oeste de la provincia, el segundo en el corazón de la provincia incluyendo la zona del GSMT, el tercero al

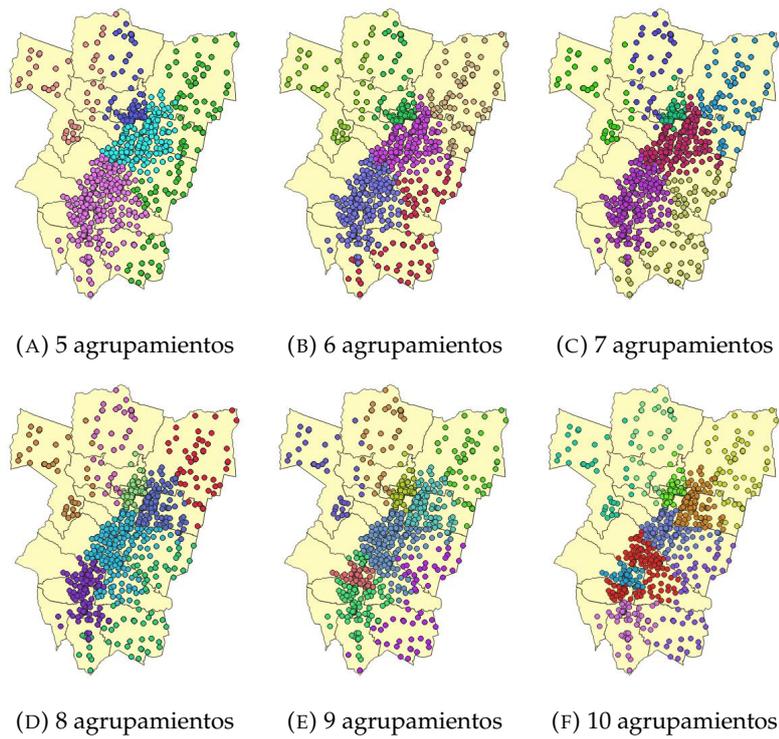


FIGURA 5.5: Resultados del método k-means donde se observa un sesgo dado por la variable departamento.

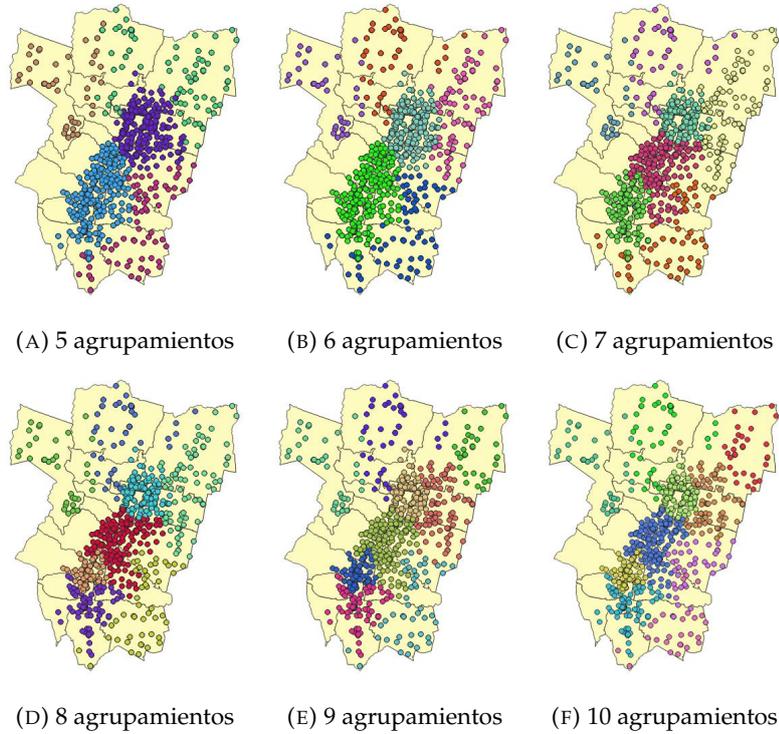


FIGURA 5.6: Resultados del método k-means con los datos puntaje total, elevacion, distancia, latitud y longitud.

sur del anterior, el mismo incluye la ciudad de Concepción, el cuarto agrupamiento se encuentra entre estos dos últimos mencionados. A medida que se incrementa la cantidad de clusters a obtener, se mantienen estos agrupamientos principales pero decreciendo en tamaño a medida que se agregan nuevos agrupamientos. Se percibe una relación entre este decrecimiento y las variables caminería y elevación, puesto que la variable distancia comienza a dominar en la aparición de nuevos agrupamientos, y la variable elevación mantiene un escalonamiento según sus valores.

5.3.2. PAM

PAM es un método de agrupamiento basado en la búsqueda de k elementos representativos (medoides). Estos elementos deben representar la estructura de los datos. Al encontrarse los medoides, se construyen k grupos, asignando cada elemento al medoide más cercano. El objetivo es minimizar la suma de las diferencias entre cada elemento y su medoide más cercano.

Siguiendo lo observado con el algoritmo K-MEANS se procedió a trabajar con dos grupos de variables, el primero comprende la totalidad de ellas, el segundo está conformado por las variables continuas descartando las categóricas para obtener resultados comparables con K-MEANS

Tomando la totalidad de variables se hicieron corridas buscando de 5 a 10 agrupamientos, siendo los resultados observados en las figuras 5.7a, 5.7b, 5.7c, 5.7d, 5.7e y 5.7f. En las mismas se puede observar que hay tres agrupamientos que casi no se modifican, estos son el agrupamiento que comprende el oeste de la provincia, y los dos agrupamientos que se encuentran en la zona pedemontana al sur de la ciudad de SMT. Sin embargo, a medida que aumenta la cantidad de agrupamientos solicitados, estos comienzan a no presentarse tan compactos como en las primeras corridas, sobre todo en la zona de la ciudad de Concepción, y en la zona este de la ciudad de SMT.

Con el grupo de variables categóricas se repitió el proceso de realizar corridas buscando de 5 a 10 agrupamientos, siendo los resultados observados en las figuras 5.8a, 5.8b, 5.8c, 5.8d, 5.8e y 5.8f. Los resultados que arroja el algoritmo tienden a mantener un eje central de agrupamiento, que se corresponde con la zona pedemontana de la provincia, desde SMT hacia el sur. Se ve claramente esta situación desde el agrupamiento siete hasta el diez, ya que se desmembran aquellos agrupamientos grandes en extensión. Esta zona pedemontana comienza disgregada en tres agrupamientos (cuando se solicitan como resultado cinco agrupamientos al algoritmo), y concluye segmentada en cinco cuando el resultado son diez agrupamientos. La zona este de la provincia se mantiene seccionada en dos agrupamientos en todos los resultados. La problemática de estos resultados se da en la zona del Río Salí, entre Simoca y Leales, el cual se presenta como una barrera natural, sin embargo en los resultados siempre hay un agrupamiento que engloba los establecimientos de la zona sur-este de la provincia. También en todos los resultados tenemos un agrupamiento que ocupa casi la totalidad del departamento Trancas y todo el norte del departamento Burruyacú, si bien estos departamento son colindantes, al querer comunicarse un edificio de Trancas con otro de Burruyacú o viceversa, deben atravesar otros agrupamientos.

5.3.3. CLARA

El método CLARA es una extensión del método PAM, para tratar con datos conteniendo un gran número de objetos, para reducir el tiempo de procesamiento y

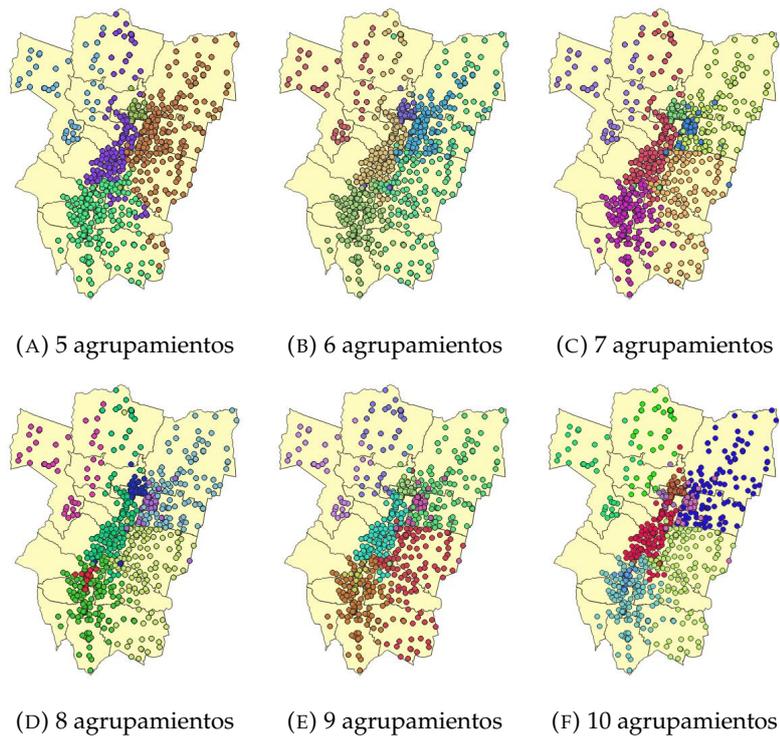


FIGURA 5.7: Resultados del método PAM.

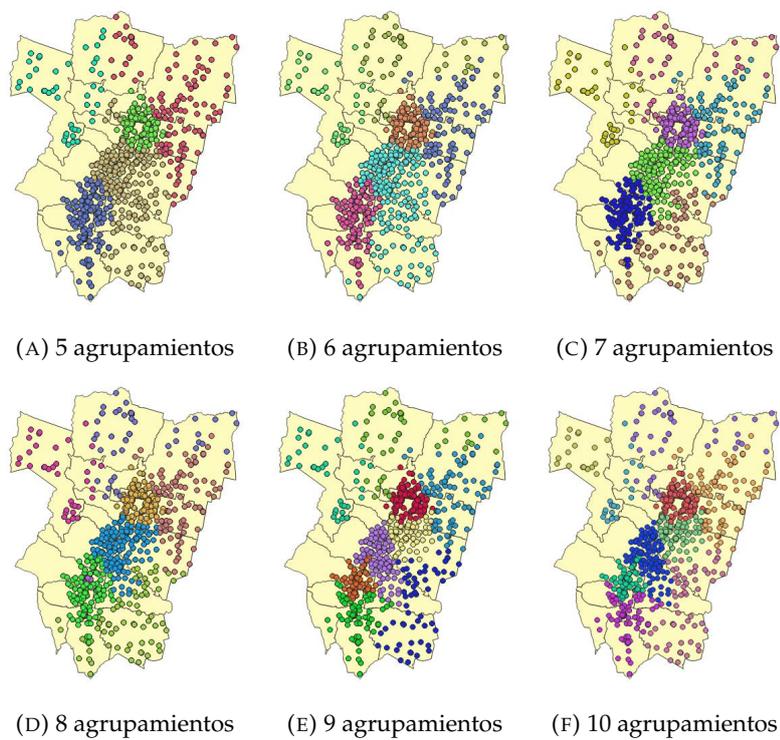


FIGURA 5.8: Resultados del método PAM.

mejorar el uso de memoria RAM, para ello utiliza métodos de muestreo.

Al correr el algoritmo para obtener de 5 a 10 agrupamientos con todas las variables, encontramos que en cada resultado tenemos agrupamientos que se dispersan

entre sí, sobretodo en la frontera entre ellos. El resultado se observa en Figuras 5.9a, 5.9b, 5.9c, 5.9d, 5.9e y 5.9f. Sin embargo se puede observar que estos clusters resultantes conforman en cada resultado tres regiones, la zona este o de llanura, la zona pedemontana o zona central de la provincia y la zona oeste (Tafí del Valle y oeste de Trancas y Tafí del Valle)

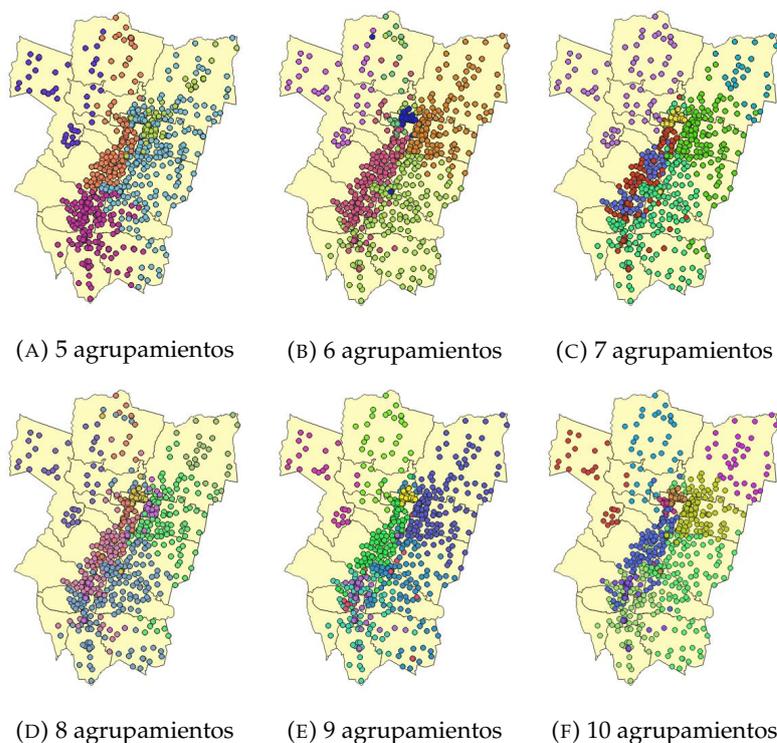


FIGURA 5.9: Resultados del método CLARA con todas las variables.

Para el caso de considerar las variables continuas, encontramos como primera diferencia que siempre se conforma un agrupamiento para la zona norte de Burruyacú y Trancas. Se mantienen tanto el corredor central de agrupamientos sobre la zona pedemontana, como la zona oeste de la provincia. Si se observa detenidamente los resultados a medida que se incrementa la cantidad de agrupamientos solicitados al algoritmo, estos tienden a conformarse alrededor de Concepción y SMT. Se observa claramente para la zona oeste la provincia, donde se mantiene a lo largo del proceso, el agrupamiento de Tafí del Valle, y oeste de Tafí Viejo y Trancas, pero en la última corrida, termina dividido en dos, cortado en sentido norte-sur, evidenciando que se divide en función de un punto central (SMT).

5.3.4. KAMILA

Otro algoritmo utilizado fue KAMILA [72]. Este método de agrupamiento equilibra equitativamente la contribución de variables continuas y categóricas. Su utilización ha permitido mejorar los resultados utilizando la totalidad de los datos con que se cuenta. Se ha trabajado sobre el conjunto de datos con el paquete RStudio, los resultados obtenidos (Figuras 5.11a, 5.11b, 5.11c, 5.11d, 5.11e y 5.11f).

Al observar los resultados parecen prometedores, podemos apreciar que los clusters obtenidos son más uniformes, encontramos una mejor partición del territorio. Sin embargo, cada vez que se ejecuta el método, con los mismos datos de entrada, los

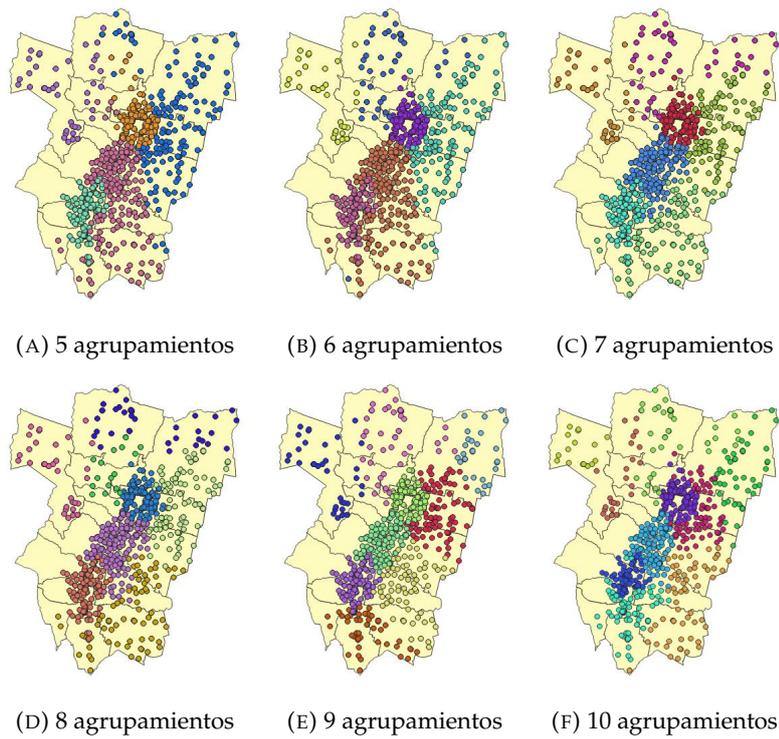


FIGURA 5.10: Resultados del método CLARA con variables continuas.

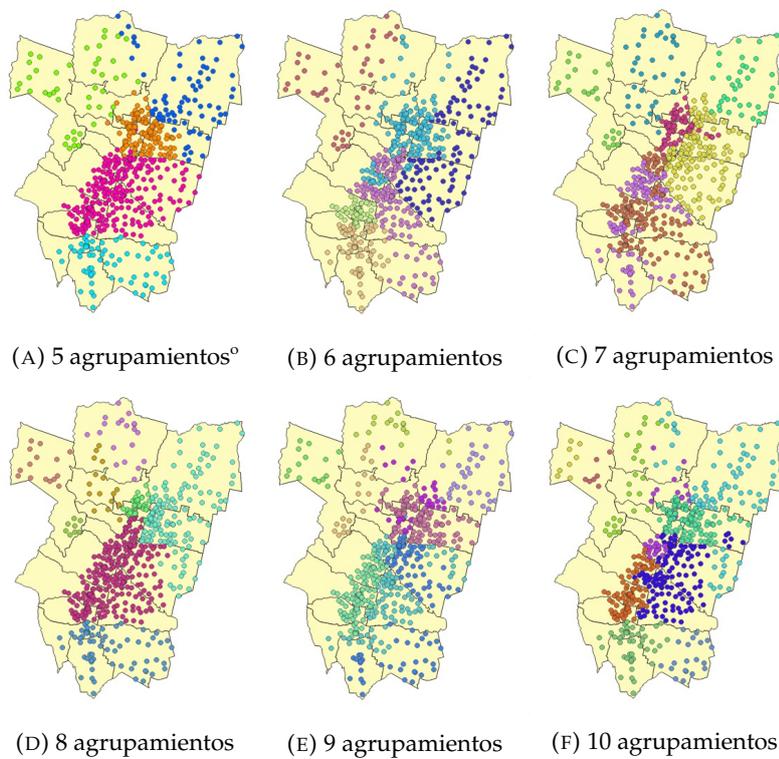


FIGURA 5.11: Resultados del método Kamila, para 5, 6, 7, 8, 9, y 10 agrupamientos.

resultados que arroja son distintos (Figuras 5.12a, 5.12b, 5.12c y 5.12d) constituyendo una de las debilidades más importantes del mencionado algoritmo. Por lo motivos mencionado, la utilización del algoritmo no representa una opción válida, dado que ante la aparición de nuevos establecimientos educativos el mismo daría resultados completamente diferentes a los anteriores.

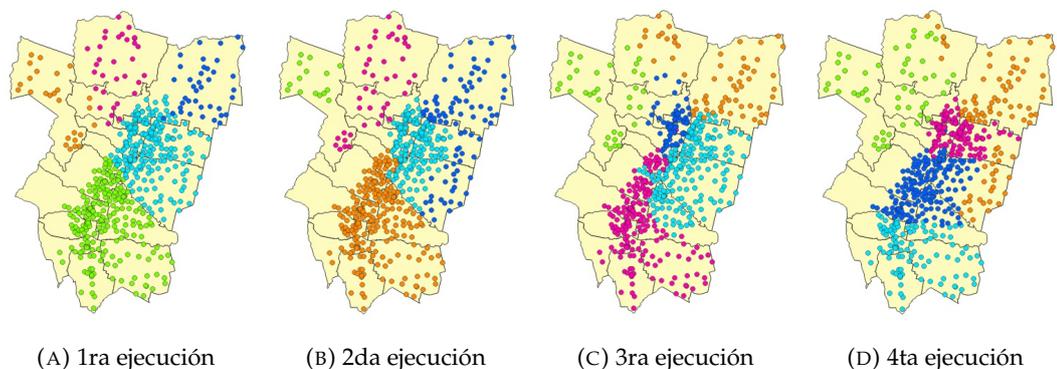


FIGURA 5.12: Resultados del método kamila sobre exactamente el mismo conjunto de datos, para obtener 5 agrupamientos.

5.4. Comparación de resultados

Al visualizar los resultados obtenidos con los algoritmos K-MEANS y CLARA podemos apreciar mucha similitud entre ellos. Sin embargo hay diferencias que a nuestro parecer inclinan la balanza hacia los resultados obtenidos por K-MEANS. Para el resultado de ocho agrupamientos, los cuales se muestran en las Figuras 5.10d y 5.6d, CLARA conforma un agrupamiento en la zona comprendida por el sur del departamento Trancas, el centro de Tafí Viejo, el oeste de Yerba Buena y centro-norte de Lules. Este cluster no aparece en los resultados de K-MEANS, pero este algoritmo configura un agrupamiento conformado en la zona comprendida por los departamentos La Cocha, Juan Bautista Alberdi, oeste de Río Chico, nor-oeste de Graneros y sur-oeste de Simoca. A nuestro criterio, para una cantidad de ocho clusters, K-MEANS conformó agrupamientos más compactos en el sur de la provincia, definiendo mejor esta zona, en cambio CLARA realizó el trabajo de división en el norte de la provincia, conformando un agrupamiento que comprende, además del descrito anteriormente, la zonas norte de los departamentos Trancas y Burruyacú.

Para el caso de nueve agrupamientos la mayor diferencia se encuentra en la zona aledaña al Río Salí en su recorrido entre Leales y Simoca, según se aprecia en Figuras 5.10e y 5.6e. En esta zona CLARA conforma un agrupamiento compacto que comprende casi la totalidad del departamento Simoca, la mitad Sur de Leales y la zona noreste de Graneros. No tiene forma de tener en cuenta que el Río es una barrera natural en esta zona. En cambio K-MEANS conforma dos agrupamientos compactos con los que divide el sur de la provincia. En la zona mencionada, K-MEANS conforma un agrupamiento pero cuya mayor concentración de elementos se encuentra al sur del río mencionado. Pudiéndose ajustar este resultado de una mejor manera que el resultado de CLARA.

5.5. Miscelaneas

Para validar los resultados obtenidos se utilizará el paquete `clValid`, que contiene tres tipos de validación: a) interna, b) de estabilidad y c) biológica. Las primeras a partir del conjunto de datos y la partición de agrupamiento utilizan información propia de los datos para evaluar la calidad del agrupamiento. Las validaciones de tipo estabilidad evalúan la consistencia de los agrupamientos obtenidos al compararlo con los grupos obtenidos al ir eliminando una variable por vez. Son de tipo biológica cuando se evalúa la capacidad de un algoritmo de agrupación para producir agrupaciones biológicamente significativas, existen medidas para investigar la homogeneidad biológica y la estabilidad de los resultados de agrupamiento.

A continuación se muestran los resultados del paquete `clValid` para el set de datos conformado por la variables continuas, los algoritmos `k-means`, `pam`, `clara` y `dbscan`. Como parámetro adicional se indica la cantidad de agrupamientos a analizar en 5, 6, 7, 8, 9 y 10 grupos.

		Clustering Methods:					
		kmeans pam clara					
		Cluster sizes:					
		5	6	7	8	9	10
		Validation Measures:					
		5	6	7	8	9	10
kmeans	APN	0.1926	0.1758	0.1240	0.2120	0.2510	0.2797
	AD	1.6613	1.5938	1.3653	1.4047	1.3193	1.2747
	ADM	0.6506	0.7138	0.3851	0.5550	0.4875	0.5743
	FOM	0.6940	0.6898	0.6066	0.6010	0.5570	0.5252
	Connectivity	93.7853	79.0742	90.3552	92.5730	98.4556	97.9456
	Dunn	0.0184	0.0265	0.0219	0.0313	0.0219	0.0245
	Silhouette	0.3941	0.4208	0.4307	0.4314	0.4265	0.3980
pam	APN	0.1720	0.2137	0.1577	0.2613	0.2780	0.2851
	AD	1.4858	1.3648	1.2006	1.2035	1.1370	1.0853
	ADM	0.5057	0.5274	0.3560	0.5036	0.4933	0.4921
	FOM	0.6762	0.6166	0.5609	0.5554	0.5396	0.4786
	Connectivity	74.7603	97.9369	135.7187	135.7187	133.8242	135.0698
	Dunn	0.0181	0.0106	0.0191	0.0191	0.0166	0.0177
	Silhouette	0.3872	0.4078	0.4285	0.4193	0.4053	0.4041
clara	APN	0.2255	0.2513	0.1756	0.2552	0.2590	0.2755
	AD	1.4976	1.4047	1.2243	1.2402	1.1659	1.0951
	ADM	0.5589	0.6032	0.4183	0.5299	0.5112	0.5070
	FOM	0.6288	0.6338	0.5618	0.5564	0.5437	0.5062
	Connectivity	105.7718	108.4917	115.2690	132.5532	136.5282	137.4421
	Dunn	0.0116	0.0123	0.0204	0.0220	0.0018	0.0039
	Silhouette	0.3711	0.3957	0.4297	0.4238	0.4302	0.4422

Para analizar estos resultados se confeccionó la Tabla 5.2, indicando en la misma para cada métrica de análisis los tres mejores resultados obtenidos. Se observa en esta tabla que el algoritmo `k-means` arroja buenos resultados en seis de las siete métricas que analiza el paquete `clValid`, si tenemos en cuenta que de 21 resultados para

las 7 métricas, 10 mejores resultados son del algoritmo k-means. También podemos observar que estos “mejores” resultados de k-means corresponden a los valores para 6, 7, 8 y 10 agrupamientos.

Por lo expuesto podemos indicar que la mayoría de los mejores resultados se corresponden al algoritmo k-means, y a su vez estos resultados se encuentran para los agrupamientos de entre 7 y 8 grupos. Correspondiéndose con el resultado analizado anteriormente.

Medida	Valores	Mejor valor	Puntuación	Método (clusters)
APN	0..1	mínimo	0,1240 0,1577 0,1720	K-means (7) Pam (7) Pam (5)
AD	0..∞	mínimo	1,0853 1,0951 1,1370	Pam (10) Clara (10) Pam (9)
ADM	0..1	mínimo	0,3560 0,3851 0,4183	Pam (7) K-means (7) Clara (7)
FOM	0..1	mínimo	0,4786 0,5062 0,5252	Pam (10) Clara (10) K-means (10)
CONNECTIVITY	0..∞	mínimo	74,7603 79,0742 90,3552	Pam (5) K-means (6) K-means (7)
DUNN	0..1	máximo	0,0313 0,0265 0,0245	K-means (8) K-means (6) K-means (10)
SILHOUETTE	-1..1	máximo	0,4422 0,4314 0,4307	Clara (10) K-means (8) K-Means (7)

CUADRO 5.2: Comparación de métricas estadísticas aplicadas a los resultados.

5.6. Conclusiones

Comenzamos este capítulo mostrando el trabajo realizado para obtener la variable "Bonificación", la cual necesita de un grupo importante de otras variables para poder ser realizado su cálculo. Comenzamos con esta variables por ser la misma importante dentro del ámbito educativo al ser utilizada para calcular el plus salarial que reciben los docentes en concepto de Zona Desfavorable.

Se continúa con la ejecución de distintos algoritmos, comenzando con el DBS-CAN, el cual, por su forma de trabajo, arrojó resultados que no se toman en cuenta. Se continuó con el algoritmo K-MEANS, este algoritmo también mostraba resultados no satisfactorios con la totalidad de variables utilizadas, pero se observó un sesgo en la composición de los resultados. Por esto, se procedió a ir eliminando del set de datos variables que podrían inducir ese sesgo. Llegando de esta forma a resultados que se muestran como prometedores, utilizando únicamente las variables de tipo continua.

A partir de lo observado con K-MEANS, se procedió a ejecutar cada algoritmo con dos set de datos, el primero conformado por la totalidad de variables (incluyendo las categóricas), el segundo con solamente las variables de tipo continuo. Se realizó esto mismo con los algoritmos PAM y CLARA.

Se comparan los resultados obtenidos, y se observa una similitud entre los resultados obtenidos con los algoritmos K-MEANS y CLARA. Se analizan en conjunto estos resultados, y se concluye que los arrojados por el algoritmo K-MEANS son los más satisfactorios.

Capítulo 6

Conclusiones y Trabajos Futuros

6.1. Conclusiones Generales

En el presente trabajo se ha propuesto una mejora en la distribución de establecimientos educativos de los circuitos educativos de la Provincia de Tucumán. Si bien, existen en la literatura numerosos trabajos sobre regionalización, no encontramos trabajos basados en la división territorial dedicados a la educación, con criterios emanados por el Ministerio de Educación de la Provincia.

Para alcanzar nuestro objetivo fue necesario definir dos grandes metas: la primera de ellas, fue la recolección de datos y su preparación para el análisis. Si bien esta información se encuentra disponible libremente en los sitios web oficiales de distintos organismos del estado, tanto nacionales como provinciales, la misma no es homogénea y no se encuentra 100 % actualizada por lo que se tuvo que trabajar para compatibilizar, acondicionar y actualizar los datos de la geolocalización de las distintas variables que se utilizó en el presente trabajo.

Debemos citar aquí el trabajo realizado para poder obtener una de las variables utilizadas, el puntaje correspondiente al concepto de Bonificación por Ubicación. Este puntaje que se otorga a cada establecimiento educativo, en función de las características de su entorno de ubicación, se calcula a partir de un grupo significativo de variables de índole espacial. La falta de precisión en la normativa que regula el cálculo del puntaje nos obligó a aplicar criterios que salven esta exigua precisión.

Para poder aplicar estos criterios se buscó información que pudiera complementar y colaborar en la definición de estos criterios. Se buscó normativa como La Ley N° 6.210 regula el Servicio de Transporte Público Automotor de Pasajeros en la Provincia de Tucumán. Esta normativa permitió aplicar definiciones en cuanto al tipo de servicio de transporte de pasajeros. La Ley Provincial N° 9.081, y su Decreto Reglamentario N° 897/7(STSV), que autorizan la continuidad de la prestación del servicio de transporte público, según lo indicado en la Ley N° 6.210. A partir de estas se pudieron dibujar y completar en un porcentaje aceptable los recorridos de las líneas de transporte de pasajeros. Esta parte del trabajo nos insumió tiempo para la búsqueda de las normativas, y el aplicar mucho de nuestro ingenio para poder resolver la situación de no poder contar con definiciones claras en la normativa que queríamos aplicar con criterios más exactos.

La segunda meta fue la redefinición de los circuitos en sí, para esto se utilizaron distintos métodos de agrupamiento largamente conocidos en la comunidad científica, como DBSCAN, KAMILA, CLARA, K-MEANS y PAM. Cada uno de los algoritmos fueron ejecutados sobre el mismo conjunto de datos, con diferentes criterios, para lograr una óptima definición de los circuitos educativos. De los mencionados, tanto CLARA como K-MEANS arrojaron resultados, que además de parecernos los más acertados, son posible de ser replicados sin mucho inconveniente. Debemos mencionar que los resultados de ambos algoritmos son parecidos, obligándonos a

compararlos con detenimiento. Dados los resultados observados con el algoritmo K-MEANS, se realizaron pruebas buscando tendencia que puedan atribuirse a algún dato en particular. Por ello se conformaron dos set de datos, el primer compuesto por la totalidad de variables, el segundo con solamente las variables de tipo continuo. Con ambos set de datos se realizaron pruebas que van desde buscar la solución con cinco clusteres hasta una máximo de diez, dado que un número mayor podría dificultar la gestión administrativa de los mismos por parte del Ministerio de Educación Provincial.

Así, consideramos que la solución está dada por la conformación de ocho clústeres realizando muy pequeños ajustes en algunos de los mismos. Si consideramos la distribución de la oferta educativa, los ocho agrupamientos obtenidos contienen cada uno todo tipo de oferta (educación inicial, primaria, secundaria y superior no universitaria), sin observarse que esta división se vea afectada por la existencia de barreras naturales (por ejemplo el Río Salí, que en su recorrido en la zona sureste de la provincia es una barrera, sin puentes o caminos que lo atraviesen). A su vez, esta división territorial puede reorganizarse agrupando estos clusters entre sí para conformar agrupamientos mayores, sin romper la lógica inicial.

6.2. Perspectiva y Trabajos Futuros

En este trabajo se propone reorganizar territorialmente los circuitos educativos de la provincia. Como continuación de este trabajo existen diversas líneas de investigación que quedan abiertas, y en las cuales es posible trabajar. Algunas directamente vinculas al objetivo de este trabajo, otras más generales pero ligadas también al territorio. Se presentan a continuación algunas líneas de trabajo futuro, entre ellas encontramos temáticas que no han podido ser tratados con la suficiente profundidad por exceder el alcance de esta tesis.

- estudiar la territorialidad de las zonas de supervisión de los distintas direcciones de nivel y proponer una mejora, para compatibilizarlas entre sí, con el objetivo de verificar que una zona de supervisión no sea solapada por más de una zona de supervisión de otra dirección de nivel.
- realizar el estudio para compatibilizar la división territorial de circuitos educativos, zonas de supervisión y la zonificación por ubicación.
- estudiar y proponer el compatibilizar la división territorial de distintos organismos (educación, salud, policía, etc.) de gobierno provincial, y así mejorar la toma de decisiones relativas a acciones conjuntas en el territorio.
- estudiar que variables pueden aplicar o dejar de aplicarse para mejorar la Resolución N° 217 del Consejo de Educación manteniendo su conceptualización pero adaptándola a los contextos sociales, económicos y productivos actuales.
- realizar el estudio de la dinámica del flujo de matrícula entre establecimientos de nivel primario a establecimientos de nivel secundario.
- estudiar la dinámica del traslado desde su residencia al establecimiento educativo por parte de docentes y alumnos.

Bibliografía

- [1] Astorga, J. Los cortes en la ruta 307 generan problemas en Taquí del Valle. *La Gaceta*, October 2011.
- [2] Dirección Nacional de Información y Estadística DiNIECE. CENPE 2014 Censo Nacional del Personal de los Establecimientos Educativos: Datos generales.
- [3] Ministerio de Educación de Tucumán Dirección de Calidad Educativa. Evolución de la matrícula, Total Provincia - Tucumán, 1999-2019. Technical report.
- [4] Dirección de Calidad Educativa Departamento de Estadística Educativa. Datos de establecimientos - Dirección de Calidad Educativa.
- [5] Infobae. El Senado le dio media sanción a la ley de Etiquetado de Alimentos. October 2020.
- [6] María Catalina Nosiglia. El proceso de sanción y el contenido de la Ley de Educación Nacional N° 26206: continuidades y rupturas. *Praxis educativa*, 11(11):113–138, 2012.
- [7] Ana Inés Heras Monner Sans and Amalia Miano. Educación, autoorganización y territorio. *Revista mexicana de investigación educativa*, 22(73):533–564, 2017.
- [8] Víctor Olaya. *Sistemas de información geográfica*. CreateSpace, 2014.
- [9] Martin Breunig, Rainer Malaka, Wolfgang Reinhardt, and Joachim Wiesel. Advancement of Geoservices. In *Information Systems in Earth Management-Kick-Off-Meeting University of Hannover*, pages 37–50, 2003.
- [10] Alejandra Beatriz Lliteras and Silvia Ethel Gordillo. Sistemas de referencias para aplicaciones SIG. In *VI Congreso Argentino de Ciencias de la Computación*, 2000.
- [11] A Marqués Mateu, J.L. Berné Valero, and J.M. Gisbert Blanquer. Transformaciones de Datum en el Contexto de un Sistema de Información Geográfica. In *EL ACCESO A LA INFORMACIÓN ESPACIAL Y LAS NUEVAS TECNOLOGÍAS GEOGRÁFICAS*, pages 959–975, Granada, España, 2006. Editorial Universidad de Granada.
- [12] Serryn Eagleson, Francisco Escobar, and IP Williamson. Hierarchical spatial reasoning applied to the automated design of administrative boundaries using GIS. 2000.
- [13] Serryn Eagleson, Francisco Escobar, and Ian Williamson. Automating the administration boundary design process using hierarchical spatial reasoning theory and geographical information systems. *International Journal of Geographical Information Science*, 17(2):99–118, 2003.

- [14] Sabine Timpf. *Hierarchical structures in map series*. Department of Geoinformation, Technical Univ., 1998.
- [15] MH Vahidnia, A Alesheikh, A Alimohammadi, and A Bassiri. Fuzzy analytical hierarchy process in GIS application. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 37(B2):593–596, 2008.
- [16] Silvana Lombardo and Massimiliano Petri. Multi Agent System: three behavioural frameworks integrated into a GIS. 2004.
- [17] Nikolaos V Karadimas, George Rigopoulos, and Nikolaos Bardis. Coupling multiagent simulation and GIS- an application in waste management. *WSEAS Transactions on Systems*, 5(10):2367–2371, 2006.
- [18] Takuto Sakamoto. Conflict Analysis in Virtual States (CAVS): A New Experimental Method Based on the Extensive Use of Multi-Agent Simulation (MAS) and Geographical Information System (GIS). 2013.
- [19] Yue Zhang and Donald E Brown. Police patrol districting method and simulation evaluation using agent-based model & GIS. *Security Informatics*, 2(1):7, 2013.
- [20] Andrew Crooks. Agent-based modeling and geographical information systems. *Geocomputation: a practical primer*. Los Angeles, London, New Delhi, Singapore, Washington DC: Sage, pages 63–77, 2015.
- [21] Jean-Francois Mas, Henri Puig, Jose Luis Palacio, and Atahualpa Sosa-Lopez. Modelling deforestation using GIS and artificial neural networks. *Environmental Modelling & Software*, 19(5):461–471, 2004.
- [22] Georgios N Kouziokas and Konstantinos Perakis. Decision support system based on artificial intelligence, GIS and remote sensing for sustainable public and judicial management. *European Journal of Sustainable Development*, 6(3):397–404, 2017.
- [23] Luca Antonio Dimuccio, Rui Ferreira, Lúcio Cunha, and António Campar de Almeida. Regional forest-fire susceptibility analysis in central Portugal using a probabilistic ratings procedure and artificial neural network weights assignment. *International Journal of Wildland Fire*, 20(6):776–791, 2011.
- [24] M Angélica Salazar Aguilar, J Luis González Velarde, and Roger Z Ríos Mercado. A divide-and-conquer approach to commercial territory design. *Computación y Sistemas*, 16(3):309–320, 2012.
- [25] Jiang Zhou and Daniel L Civco. Using genetic learning neural networks for spatial decision making in GIS. *Photogrammetric Engineering and Remote Sensing*, 62(11):1287–1295, 1996.
- [26] Steven Van Dijk, Dirk Thierens, and Mark De Berg. Using genetic algorithms for solving hard problems in GIS. *GeoInformatica*, 6(4):381–413, 2002.
- [27] Aissa Taibi and Baghdad Atmani. Combining Fuzzy AHP with GIS and Decision Rules for Industrial Site Selection. *International Journal of Interactive Multimedia & Artificial Intelligence*, 4(6), 2017.

- [28] *OpenGIS Geography Markup Language (GML) Encoding Standard 3.2.1*. Number 07-036. OGC, 2007.
- [29] *OpenGIS Web Coverage Service (WCS) Implementation Specification 1.1.2*. Number 07-067r5. OGC, 2008.
- [30] *OpenGIS Web Map Service (WMS) Implementation Specification 1.3.0*. Number 06-042. OGC, 2006.
- [31] *Web Feature Service (WFS) Implementation Specification 1.1.0*. Number 04-094. OGC, 2005.
- [32] Ministerio de Educación. *Resolución N° 753/5 MEd*. September 2009.
- [33] Ministerio de Educación. *Resolución N° 1070/5 MEd*. October 2011.
- [34] Superior Gobierno de la Provincia de Tucumán. *La escuela, un espacio para reponer las palabras*. Department of Geoinformation, Technical Univ., 2015.
- [35] *Repavimentación y Mejoramiento RP 307: Tramo Acheral - Tafí del Valle Evaluación de Impacto Ambiental*.
- [36] M. Rosa Brumat. Maestros rurales: condiciones de trabajo, formación docente y práctica cotidiana. *Revista iberoamericana de educación*, 55(4):1–10, 2011.
- [37] Patricia Monica Fernández and Daniel Esteban Quiroga. Aportes metodológico-conceptuales de la distribución espacial de la población: El caso de Tucumán y Catamarca en el Noroeste Argentino [1991, 2001 y 2010]. In *VII Jornadas de Sociología de la UNLP 5 al 7 de diciembre de 2012 La Plata, Argentina. Argentina en el escenario latinoamericano actual: Debates desde las ciencias sociales*. Universidad Nacional de La Plata. Facultad de Humanidades y Ciencias de la ..., 2012.
- [38] María E Puchulu and Diego S Fernández. Características y distribución espacial de los suelos de la provincia de Tucumán. *Geología de Tucumán'*.(Eds S Moyano, ME Puchulu, D Fernandez, G Aceñolasa, ME Vides, S Nieva) pp, pages 1–17, 2016.
- [39] Matilde Malizia and Paula Boldrini. Las lógicas de ocupación del espacio urbano. Un estudio de realidades contrapuestas. El caso de las urbanizaciones cerradas y villas miseria en Yerba Buena, Gran San Miguel de Tucumán. *Cuadernos de la Facultad de Humanidades y Ciencias Sociales-Universidad Nacional de Jujuy*, (41):197–219, 2012. Publisher: Universidad Nacional de Jujuy.
- [40] Consejo de Educación. *Resolución N° 217*, March 1991.
- [41] Pilar Moreno, Jesús García, and LDE Lacalle. Estado del arte en procesos de zonificación. *GeoFocus*, (11):155–181, 2011.
- [42] Cristina Garcia Cambronero and Irene Gómez Moreno. Algoritmos de aprendizaje: knn & kmeans. *Inteligencia en Redes de Comunicación, Universidad Carlos III de Madrid*, 23, 2006.
- [43] Isabel Cristina Pérez Verona and Leticia Arco García. Una revisión sobre aprendizaje no supervisado de métricas de distancia. *Revista Cubana de Ciencias Informáticas*, 10(4):43–67, 2016. Publisher: Universidad de las Ciencias Informáticas.

- [44] Claude Sammut and Geoffrey I Webb. *Encyclopedia of machine learning*. Springer Science & Business Media, 2011.
- [45] Alicia Inés Ferrari, Agustín Jose Machuca, and Sandra L Mansilla. Transformación y articulación del territorio urbano en el Gran San Miguel de Tucumán: El caso de Yerba Buena. *Breves Contribuciones del Instituto de Estudios Geográficos*, (27):89–105, 2016. Publisher: Instituto de Estudios Geográficos "Dr. Guillermo Rohmeder". Facultad de . . .
- [46] Hada Melissa Sáenz Vela. Revisando los métodos de agregación de unidades espaciales: MAUP, algoritmos y un breve ejemplo. *Estudios demográficos y urbanos*, 31(2):385–411, 2016. Publisher: El Colegio de México.
- [47] Santiago Garganta. *El Impacto laboral y demográfico de la Asignación Universal por hijo en Argentina*. PhD Thesis, Facultad de Ciencias Económicas, 2015.
- [48] Axel Rivas, Alejandro Vera, and Pablo Bezem. *Radiografía de la educación argentina*. Cippec, 2010.
- [49] Roberto Mur Montero. Notas sobre la organización desconcentrada de las administraciones educativas. *Revista Iberoamericana de Educación*, (3):41–62, 1993.
- [50] Silvio Cardona González. Educación, Territorio y Modernización en el Contexto de la Descentralización del Sistema Educativo en Colombia en Orden al Territorio. *Revista Ideales*, 1(1), 2016.
- [51] Rodolfo Montaña Salazar and Gerardo Núñez Medina. Metodología para medir la concentración y dispersión demográfica de la población mexicana a través del modelo policéntrico.
- [52] Esther García Cuadrado. Los sistemas de información geográfica como recurso didáctico en la enseñanza de la Biología y Geología en Secundaria. 2013.
- [53] Luis Eduardo Oliveira Lizardo and Clodoveu Augusto Davis Jr. A PostGIS extension to support advanced spatial data types and integrity constraints. In *Proceedings of the 25th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pages 1–10, 2017.
- [54] José Ramón Rodríguez Pérez, Xesús Pablo González Vázquez, and Pedro Arias Sánchez. Cartografía de usos del suelo por fotointerpretación mediante Sistemas de Información Geográfica (SIG): análisis comparativo de los sistemas ráster y vectorial. *Actas del XI Congreso Internacional de Ingeniería Gráfica*, 3:1513–1523, 1999.
- [55] Felix Bensmann, Rainer Roosmann, Roman Wössner, Andreas Abecker, and Carsten Heidmann. RichWPS Orchestration Environment for Geo Services. In *ICHE 2014. Proceedings of the 11th International Conference on Hydrosience & Engineering*, pages 1077–1084, 2014.
- [56] M Bitar. *La política de descentralización educativa. ciudadanía social e impacto en la administración pública y en el entramado social*. PhD Thesis, Tesis de maestría). Universidad Católica de Santa Fe. Argentina, 1999.
- [57] José Antonio Remis, Eduardo Alfredo Guarnera, and Alberto Parra. Impacto de la hidatidosis. Influencia de factores ambientales y socioculturales en Tucumán, Argentina. *Revista argentina de radiología*, 73(3):303–312, 2009. Publisher: Sociedad Argentina de Radiología.

- [58] Jesús García Herrero and José Manuel Molina López. Técnicas de análisis de datos. 2006.
- [59] Miguel Ángel Manso Callejo. Metadatos en los sistemas de información geográfica (ISO-19115). 2003.
- [60] Konstantinos Evengelidis, Anastasios Agrianidis, Konstantinos Perakis, Theofilos Papadopoulos, and Konstantinos Papatheodorou. Web-GIS Development of Geospatial Data Dissemination in EU Operational Programmes. *European Journal of Geography*, 9(2):21–36, 2018.
- [61] Diana Comesaña and Bryan Barreiro. Metadatos Geográficos: El perfil Uruguayo.
- [62] Douglas Nebert, Arliss Whiteside, and P Vretanos. OGC catalogue services specification. *Open Geospatial Consortium Inc*, 2005.
- [63] Instituto Geográfico Nacional. Diccionario de Conceptos y Catálogo de Objetos Geográficos del Instituto Geográfico Nacional (ISO/TC211 19110 y 19126 - DFDD), November 2017.
- [64] Gabriel Atilio Rivas. Métodos de análisis espacial aplicados al estudio de la agricultura intensiva en el Partido de La Plata. In *VII Jornadas de Investigación del Centro de Investigaciones Geográficas y del Departamento de Geografía 8 de noviembre de 2005 La Plata, Argentina*. Universidad Nacional de La Plata. Facultad de Humanidades y Ciencias de la . . . , 2005.
- [65] Mariella Gutiérrez. El rol de las bases de datos espaciales en una infraestructura de datos. In *GSDI-9 Conference Proceedings*, pages 6–10, 2006.
- [66] Axel J Soto, Ignacio Ponzoni, and Gustavo E Vazquez. Análisis Numérico De Diferentes Criterios De Similitud En Algoritmos De Clustering. *Mecánica Computacional*, pages 993–1012, 2006.
- [67] María Ximena Dueñas Reyes. Minería de datos espaciales en búsqueda de la verdadera información. *Ingeniería y universidad*, 13(1):7, 2009. Publisher: Facultad de Ingeniería.
- [68] Man Lung Yiu and Nikos Mamoulis. Clustering objects on a spatial network. In *Proceedings of the 2004 ACM SIGMOD international conference on Management of data*, pages 443–454, 2004.
- [69] Ramón Giraldo. Análisis exploratorio de variables regionalizadas con métodos funcionales. *Revista Colombiana de Estadística*, 30(1):115–127, 2007.
- [70] D Pascual, F Pla, and S Sánchez. Algoritmos de agrupamiento. *Método Informáticos Avanzados*, pages 164–174, 2007.
- [71] Alex Foss, Marianthi Markatou, Bonnie Ray, and Aliza Heching. A semiparametric method for clustering mixed data. *Machine Learning*, 105(3):419–458, 2016. Publisher: Springer.
- [72] Alexander H Foss and Marianthi Markatou. kamila: clustering mixed-type data in R and Hadoop. *Journal of Statistical Software*, 83(1):1–44, 2018.
- [73] Jiawei Han, Micheline Kamber, and Jian Pei. *Data mining: concepts and techniques*. Elsevier, 2011.

- [74] Abdelhalim Bendib. The effects of spatial clustering of public facilities on social equity and urban congestion in the city of Batna (Algeria). *GeoJournal*, pages 1–14, 2020. Publisher: Springer.
- [75] Mohamed Elamin Ahmed Babiker and Hajir Khaïd Elshaikh Idris. GIS Based Services Using Network Analysis of Khartoum North.
- [76] Lorien Nesbitt, Michael J Meitner, Cynthia Girling, Stephen RJ Sheppard, and Yuhao Lu. Who has access to urban vegetation? A spatial analysis of distributional green equity in 10 US cities. *Landscape and Urban Planning*, 181:51–79, 2019. Publisher: Elsevier.
- [77] Jack Cordes and Marcia C Castro. Spatial analysis of COVID-19 clusters and contextual factors in New York City. *Spatial and Spatio-temporal Epidemiology*, 34:100355, 2020. Publisher: Elsevier.
- [78] Czesław Adamiak, Barbara Szyda, Anna Dubownik, and David García-Álvarez. Airbnb offer in Spain—spatial analysis of the pattern and determinants of its distribution. *ISPRS international journal of geo-information*, 8(3):155, 2019. Publisher: Multidisciplinary Digital Publishing Institute.
- [79] Yadian Guillermo Pérez Betancourt, Liset González Polanco, Rolando Morales Pérez, and Yanislay Torres Vega. Estratificación de territorios basada en indicadores de salud sobre el Sistema de Información Geográfica QGIS. *Revista Cubana de Ciencias Informáticas*, 10:163–175, 2016. Publisher: Universidad de las Ciencias Informáticas.
- [80] Josep Mercadé Aloy, Francesc Magrinyà Torner, and Marina Cervera Alonso de Medina. Descifrando la forma urbana: un análisis de patrones de agrupamiento basado en SIG. *Geofocus: revista internacional de ciencia y tecnología de la informacion geográfica*, (22):3–19, 2018.
- [81] Juan Manuel Suárez Rodríguez. Caracterización de los Hurtos a Personas que Afectan la Localidad Los Mártires de la Ciudad de Bogotá, Mediante la implementación de Algoritmos de Agrupamiento de Minería de Datos Espaciales y Apoyado en una Infraestructura de Datos Espacial.
- [82] Anastasia Rava and Claudia Santa Ana. Desarrollo de un módulo de generación de zonas de manejo para un Sistema de Información Geográfico (SIG) de Gestión Agrícola. 2008. Publisher: UR. FI-INCO,.
- [83] Hee Jeong Yun, Dong Jin Kang, Dong-Kap Kim, and Youngeun Kang. A GIS-Assisted Assessment and Attribute-Based Clustering of Forest Wetland Utility in South Korea. *Sustainability*, 11(17):4632, 2019. Publisher: Multidisciplinary Digital Publishing Institute.
- [84] Edward Boampong and K. Mert Çubukçu. The Organization of Urban Space and Socio-Economic Characteristics: A Graph Theory-Based Empirical Study Using Hierarchical Cluster Analysis Kentsel Mekân Organizasyonu ve Sosyo-Ekonomik Yapı.
- [85] Michael S Rosenberg and Corey Devin Anderson. PASSaGE: pattern analysis, spatial statistics and geographic exegesis. Version 2. *Methods in Ecology and Evolution*, 2(3):229–232, 2011. Publisher: Wiley Online Library.

- [86] Michael S Rosenberg. The bearing correlogram: a new method of analyzing directional spatial autocorrelation. *Geographical Analysis*, 32(3):267–278, 2000. Publisher: Wiley Online Library.
- [87] Michael S Rosenberg. Wavelet analysis for detecting anisotropy in point patterns. *Journal of Vegetation Science*, 15(2):277–284, 2004. Publisher: Wiley Online Library.
- [88] Simon Schmidt, Christine Alewell, Panos Panagos, and Katrin Meusburger. Regionalization of monthly rainfall erosivity patterns in Switzerland. *Hydrology and Earth System Sciences*, 20(10):4359–4373, 2016. Publisher: Copernicus.
- [89] Yadian Guillermo Pérez Betancourt, Liset González Polanco, Juan Pedro Febles Rodríguez, and Alcides Cabrera Campos. Propuestas para el análisis geoespacial en estudios salubristas. *Revista Cubana de Ciencias Informáticas*, 12(2):44–57, 2018. Publisher: Universidad de las Ciencias Informáticas.
- [90] Tomáš Středa, Radim Cerkal, Lenka Hájková, Filip Chuchma, Tomáš Khel, and Jana Klimešová. GIS application in abiotic risks regionalization for spring barley. *Contributions to Geophysics and Geodesy*, 50(1):49–60, 2020.
- [91] Vincent Ezikornwor Weli, Alexander Chibuzor Okoli, and Shedrach Worlu. The Algorithms of Urban Sprawl Dynamics on Surface Temperature Characteristics of Greater Port Harcourt Region, Using Remote Sensing-GIS Approach. *Annals of Geographical Studies*, 3(2):19–33, 2020.
- [92] Claudia Baxendale and Gustavo Buzai. Caracterización socioespacial de la cuenca del río Luján (Provincia de Buenos Aires, Argentina). *Memorias Universidad del Azuay*, (XVI):253–259, 2017.
- [93] Peter Lentjes, Michael Peters, and Federico Holmann. Regionalization of climatic factors and income indicators for milk production in Honduras. *Ecological Economics*, 69(3):539–552, 2010. Publisher: Elsevier.
- [94] Armando Avalos Jimenez, Fernando Florez Vilchez, Susana Marcelleño Flores, María Lorena, and others. Integración del paisaje como propuesta de regionalización. Cuenca río Mololoa. CONACYT, 2015. Publisher: Revista Iberoamericana de las Ciencias Biológicas Agropecuarias.
- [95] Fahui Wang, Cuiling Liu, and Yaping Xu. Analyzing population density disparity in china with GIS-automated regionalization: The hu line revisited. *Chinese Geographical Science*, 29(4):541–552, 2019. Publisher: Springer.
- [96] Farhad Zolfaghari, Hassan Khosravi, Alireza Shahriyari, Mitra Jabbari, and Azam Abolhasani. Hierarchical cluster analysis to identify the homogeneous desertification management units. *Plos one*, 14(12):e0226355, 2019. Publisher: Public Library of Science San Francisco, CA USA.
- [97] Vanina Solange Aliaga, Federico Ferrelli, ED Alberdi Algorañaz, Vanesa Yael Bohn, and Maria Cintia Piccolo. Distribution and variability of precipitation in the Pampas, Argentina. *Cuadernos de investigación Geográfica*, 42(1):261–280, 2016.
- [98] Adalto Moreira Braz, Ivanilton José de Oliveira, Lucas Costa de Souza Cavalcanti, António Campar de Almeida, and Eduardo Salinas Chávez. Cluster

- Analysis For Landscape Typology. *Mercator (Fortaleza)*, 19, 2020. Publisher: SciELO Brasil.
- [99] N° Ley. 26206/2006. Ley de Educación Nacional. *Boletín Oficial. Argentina*, 28.
- [100] M Criado, M Crespo, C Rodríguez, M Bravo, and D Ballari. Creación de Metadatos: Metodología y experiencia del Grupo de Catalogadores de la Información Geográfica. *Jornadas de las Infraestructuras de Datos Espaciales de España*, 2007.
- [101] Carlos Mena Frau, Yony Ormazábal Rojas, Yohana Morales Hernández, and John Gajardo Valenzuela. Exactitud espacial en la creación de bases de datos SIG modelos raster y vectorial. *Ingeniare. Revista chilena de ingeniería*, 16(1):159–168, 2008. Publisher: Universidad de Tarapacá.
- [102] Jean-François Mas and Tania Fernandez. Una evaluación cuantitativa de los errores en el monitoreo de los cambios de cobertura por comparación de mapas. *Investigaciones geográficas*, (51):73–87, 2003. Publisher: Instituto de Geografía, UNAM.
- [103] Vidala del Rosario Cáceres De Gill, Nilsa Elynor Gill Cáceres, and María Zunilda Galeano Benítez. Incertidumbre docente por Pandemia COVID-19. *Academic Disclosure*, 1(1):96–108, 2020.
- [104] José Luis Arocha Reyes. La Geografía y la Cartografía: Dos disciplinas inseparables. *Universidad Central de Venezuela. Facultad de Humanidades y Educación. Caracas*, 1985.
- [105] João Francisco de Abreu. First law of geography, first computer map, cartograms, analytical cartography, wind model—tribute to Waldo Tobler: scientist, geographer, cartographer. *Annals of GIS*, 24(3):221–222, 2018. Publisher: Taylor & Francis.
- [106] Andrés Mauricio Basto Prieto. Análisis espacial para el estudio de una posible invasión mecanizada del Ejército Venezolano en la Orinoquia colombiana para la ocupación de la ciudad de Villavicencio. Publisher: Universidad Militar Nueva Granada.
- [107] Adriana Concepción Ceballos Montenegro. Análisis espacial aplicado a minas antipersona en el departamento de Antioquia. Publisher: Universidad Militar Nueva Granada.
- [108] Cynthia Ruth Matusevich. Análisis espacial con SIG orientados a la obtención de sitios candidatos para la localización de la disposición final de residuos sólidos urbanos (RSU) en la ciudad de Corrientes. 2016. Publisher: Universidad Nacional del Nordeste. Facultad de Humanidades.
- [109] Carl Reed, Kurt Buehler, and Lance McKee. OGC consensus: How successful standards are made. *ISPRS International Journal of Geo-Information*, 4(3):1693–1706, 2015. Publisher: Multidisciplinary Digital Publishing Institute.
- [110] Willington Siabato. Sobre la evolución de la información geográfica: las bodas de oro de los SIG. *Cuadernos de Geografía: Revista Colombiana de Geografía*, 27(1):1–9, 2018. Publisher: Facultad de Ciencias Humanas de la Universidad Nacional de Colombia.

- [111] Laura García Juan and Alejandro Vallina Rodríguez. SIG y bases de datos. Oportunidades y retos en la transición de los sistemas tradicionales al Big Data= GIS and databases. Opportunities and Challenges in the Transition from Traditional Systems to Big Data. *Espacio Tiempo y Forma. Serie VI, Geografía*, (12):135–158, 2019.
- [112] Laura García Juan, Alejandro Vallina Rodríguez, and Carmen Hidalgo Giralt. De los SIG a las TIG: El encaje de sistemas tradicionales en la sociedad del Big Data y de la IDE.
- [113] Laura García Juan and Ángel Ignacio Aguilar Cuesta. Big Data, una tecnología de hoy para el análisis del pasado. In *Naturaleza, territorio y ciudad en un mundo global*, pages 2594–2603. Asociación de Geógrafos Españoles, 2017.
- [114] J Gentili and Verónica Gil. Aplicación de Tecnologías de Información Geográficas (TIG) al análisis hidrogeomorfológico. *Caso de estudio: cuencas hidrográficas del Sistema de Ventania. Buenos Aires. Argentina*, 12, 2009.
- [115] Javier Gutiérrez Puebla. Big Data y nuevas geografías: la huella digital de las actividades humanas. *Documents d'anàlisi geogràfica*, 64(2):0195–217, 2018.
- [116] FJ Teo. SharMap software libre para aplicaciones SIG. *Mapping*,(188), 28, 34, 2018.
- [117] Rubén Olea and Pedro Araujo. Aplicación de los Sistemas de Información Geográficas (SIG) en el análisis de la bonificación por ubicación. In *49 Jornadas Argentinas de Informática JAIIO - Simposio de Informática en el Estado*, 2020.
- [118] Francisco R Feito Higuera and Rafael J Segura Sánchez. Herramientas SIG 3D. *Virtual Archaeology Review*, 1(1):87–91, 2010. Publisher: Universitat Politècnica de València.
- [119] José Antonio Jiménez Berni, María Aguilera Urueña, and J Meroño de Larriva. Alternativas de software libre a los sistemas de información geográfica comerciales. *Sevilla, España. Obtenido de <http://www.cartesia.org/geodoc/ingegraf2005/gis10.pdf>*, 2005.
- [120] Castro Benavides, Brian Castro, and others. Obtención de una cuenca hidrográfica mediante el uso de un software comercial “Arcgis” y el uso de un software libre “Saga”. Caso de estudio municipio de Porcesito, Antioquia. Publisher: Universidad Militar Nueva Granada.
- [121] Carlos Guerra Flores. Implementación de un geoportal utilizando mapserver y pmapper para la empresa celepsa empleando software libre. 2019. Publisher: Universidad Nacional Federico Villarreal.
- [122] Alberto Miguel Rodrigo Martínez. Migration of ESRI server data to an Apache server with Geoserver and Openlayers. Settlement development in Vel'ké Pole, Pila and Radobica, Slovakia. 2020.
- [123] David Bucklin and Mathieu Basille. Rpostgis: Linking R with a PostGIS Spatial Database. *R Journal*, 10(1), 2018.
- [124] Hyung-Gyu Ryoo, Soojin Kim, Joon-Seok Kim, and Ki-Joune Li. Development of an extension of GeoServer for handling 3D spatial data. In *Free and Open Source Software for Geospatial (FOSS4G) Conference Proceedings*, volume 17, page 6, 2017. Issue: 1.

- [125] Anil K Jain. Data clustering: 50 years beyond K-means. *Pattern recognition letters*, 31(8):651–666, 2010. Publisher: Elsevier.
- [126] Martin Ester, Hans-Peter Kriegel, Jörg Sander, Xiaowei Xu, and others. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Kdd*, volume 96, pages 226–231, 1996. Issue: 34.
- [127] Leonard Kaufman and Peter J. Rousseeuw. *Clustering by means of medoids*. 1987.
- [128] Leonard Kaufman and Peter Rousseeuw. *Clustering Large Data Sets*. pages 425–437. 1986.
- [129] Raymond T. Ng and Jiawei Han. *Efficient and Effective Clustering Methods for Spatial Data Mining*. 1994.
- [130] Michael F Goodchild and Robert P Haining. SIG y análisis espacial de datos: perspectivas convergentes. *Investigaciones Regionales-Journal of Regional Research*, (6):175–201, 2005. Publisher: Asociación Española de Ciencia Regional.
- [131] Grethell Castillo Reyes. Técnicas de programación paralela aplicadas al procesamiento de datos ráster mediante la biblioteca GDAL. *Revista Cubana de Ciencias Informáticas*, 10(1):71–94, 2016. Publisher: Universidad de las Ciencias Informáticas.
- [132] Javier Valencia Martínez de Antoñana. *Pasado, presente y futuro de las infraestructuras de datos espaciales*. 2013.
- [133] José María Luna-Romera, Jorge García-Gutiérrez, María Martínez-Ballesteros, and José C Riquelme Santos. Aproximación al índice externo de validación de clustering basado en chi cuadrado.
- [134] Peter J Rousseeuw. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, 20:53–65, 1987.
- [135] Guy Brock, Vasyl Pihur, Susmita Datta, Somnath Datta, and others. clValid, an R package for cluster validation. *Journal of Statistical Software (Brock et al., March 2008)*, 2011.
- [136] ES Samundeeswari and M Kiruthika. Clustering Similar Images Using Various Image. *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)*, 7(3), 2018.
- [137] Ka Yee Yeung, David R. Haynor, and Walter L. Ruzzo. Validating clustering for gene expression data. *Bioinformatics*, 17(4):309–318, 2001. Publisher: Oxford University Press.
- [138] Joseph C Dunn. Well-separated clusters and optimal fuzzy partitions. *Journal of cybernetics*, 4(1):95–104, 1974. Publisher: Taylor & Francis.
- [139] Anand Bihari, Sudhakar Tripathi, and Akshay Deepak. Gene Expression Analysis Using Clustering Techniques and Evaluation Indices. In *Proceedings of 2nd International Conference on Advanced Computing and Software Engineering (ICACSE)*, 2019.

-
- [140] Peter J Rousseeuw. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, 20:53–65, 1987. Publisher: Elsevier.
- [141] Chunhui Yuan and Haitao Yang. Research on K-value selection method of K-means clustering algorithm. *J—Multidisciplinary Scientific Journal*, 2(2):226–235, 2019. Publisher: Multidisciplinary Digital Publishing Institute.
- [142] Abba Chouni Benabdellah, Asmaa Benghabrit, and Imane Bouhaddou. A survey of clustering algorithms for an industrial context. *Procedia computer science*, 148:291–302, 2019. Publisher: Elsevier.