

Trabajo Fin de Máster

PROPUESTA DE ESTRATEGIAS PARA LA MODELACIÓN DE UN SISTEMA DE ABASTECIMIENTO CON INFORMACIÓN LIMITADA. CASO DE ESTUDIO DEL MUNICIPIO DE NOGOYÁ (ENTRE RÍOS, ARGENTINA).

Intensificación: *HIDRÁULICA URBANA*

Autor:

LAURA PASTOR JABALOYES

Director:

DR. RAFAEL PÉREZ GARCÍA

Codirector/es:

JOAQUÍN IZQUIERDO SEBASTIÁN



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA

SEPTIEMBRE, 2015

máster en ingeniería
hidráulica y medio ambiente

mihma



Resumen del Trabajo de Fin de Máster

Datos del proyecto

Título: Propuesta de estrategias para la modelación de un sistema de abastecimiento con información limitada. Caso de estudio del municipio de Nogoyá (Entre Ríos, Argentina).

Alumno: Laura Pastor Jabaloyes.

Director: Dr. Rafael Pérez García

Codirector/es:

Dr. Joaquín Izquierdo Sebastián

Fecha de Lectura: Septiembre, 2015

Resumen

En castellano (máximo 2000 palabras)

En la actualidad, sobre todo en países en vías de desarrollo, un gran número de medianas y pequeñas poblaciones se enfrentan al problema del aumento de la demanda de agua potable, si bien, no disponen de herramientas y medios para dar una solución fundamentada desde el punto de vista técnico. El empleo de modelos estratégicos de la red de suministro sería, para dichas poblaciones, un potente soporte en la toma de decisiones, particularmente dentro del ámbito de la planificación de esta infraestructura en el medio y largo plazo. No obstante, el contexto en el que se presenta esta necesidad suele estar caracterizado por la escasez, entre otros no menos importantes, de información básica para la aplicación de técnicas de modelación convencionales, sumada a una situación sociopolítica y económica que no permite paliar esta carencia en el corto plazo a través de los recursos técnicos habituales. En consecuencia, la pregunta que surge es: ¿cómo conseguir un modelo matemático de una red de abastecimiento de agua bajo estas condiciones, que arroje resultados útiles para la toma de decisión en la planificación estratégica que afectan a esta infraestructura?

El objetivo principal de este trabajo es dar respuesta a la pregunta planteada. Para ello, se pretende complementar la metodología convencional de modelación con el aprovechamiento de información

secundaria y procedente de poblaciones distintas a la del problema a resolver, esencialmente a través de técnicas de clasificación, como el Análisis Clústering y los Support Vector Machines. En la fase de calibración, se pretende absorber la incertidumbre de una realidad que dista mucho de estar descrita de forma clara y unívoca con la asistencia de técnicas y conceptos propios del Soft Computing: los algoritmos genéticos y las funciones de pertenencia de los subconjuntos difusos.

El presente proyecto parte del caso de estudio del municipio de Nogoyá (Argentina), escogido a raíz de una colaboración con el Centro de Estudios para la Gestión Local Sostenible del Agua y del Hábitat Humano, perteneciente a la Universidad Autónoma de Entre Ríos, posible gracias a una beca del Centro de Cooperación al Desarrollo de la Universitat Politècnica de València.

El caso de Nogoyá presenta particularidades que requieren de la consecución de una serie de objetivos parciales. Los mismos quedan circunscritos principalmente a recabar toda la información que esté disponible (este objetivo se cumplió casi en su totalidad durante los 3 meses de estancia en Paraná (Entre Ríos)) y procesarla para que quede almacenada en un formato funcional. Para esto último se inicia un proyecto en un Sistema de Información Geográfica, con el objetivo de crear un repositorio de gran potencial con los datos disponibles sobre la red de abastecimiento de agua de Nogoyá, en tanto que dichos datos quedan georeferenciados.

En valenciano (máximo 2000 palabras)

En l'actualitat, sobretot en països en vies de desenvolupament, una gran quantitat de mitjanes i petites poblacions s'enfronten al problema de l'augment de la demanda d'aigua potable, si bé, no disposen d'eines i mitjans per a donar una solució fonamentada des del punt de vista tècnic. L'ús de models estratègics de la xarxa de subministrament seria, per a aquestes poblacions, un potent suport a la toma de decisions, particularment dins de l'àmbit de la planificació d'aquesta infraestructura en el mitjà i llarg termini. No obstant açò, el context en el qual es presenta aquesta necessitat sol estar caracteritzat per l'escassetat, entre uns altres no menys importants, d'informació bàsica per a l'aplicació de tècniques de modelació convencionals, sumada a una situació sociopolítica i econòmica que no permet pal·liar aquesta falta en el curt termini a través dels recursos tècnics habituals. En conseqüència, la pregunta que sorgeix és: com aconseguir un model matemàtic d'una xarxa de proveïment d'aigua sotmesa a aquestes condicions, que obtinga resultats útils per a la toma de decisió en la planificació estratègica que afecten a aquesta infraestructura?

L'objectiu principal d'aquest treball és donar resposta a la pregunta plantejada. Per dur a terme açò, es pretén complementar la metodologia convencional de modelació amb l'aprofitament d'informació secundària i procedent de poblacions diferents a la del problema a resoldre, essencialment a través de tècniques de classificació, com l'Anàlisi Clústering i els Support Vector Machines. En la fase de calibratge, es pretén absorbir la incertesa d'una realitat que dista molt d'estar descrita de forma clara i unívoca amb l'assistència de tècniques i conceptes propis del Soft Computing: els algorismes genètics i les funcions de pertinença dels subconjunts difusos.

El present projecte part del cas d'estudi del municipi de Nogoyá (Argentina), escollit arran d'una col·laboració amb el Centro de Estudios para la Gestión Local Sostenible del Agua y del Hábitat Humano,

pertanyent a la Universidad Autónoma de Entre Ríos, possible gràcies a una beca del Centro de Cooperación al Desarrollo de la Universitat Politècnica de València.

El cas de Nogoyá presenta particularitats que requereixen de la consecució d'una sèrie d'objectius parcials. Els mateixos queden circumscrits principalment a recaptar tota la informació que estiga disponible (aquest objectiu es va complir quasi íntegrament durant els 3 mesos d'estada en Paraná (Entre Ríos)) i processar-la perquè quede emmagatzemada en un format funcional. Per a açò últim s'inicia un projecte en un Sistema d'Informació Geogràfica, amb l'objectiu de crear un repositori de gran potencial amb les dades disponibles sobre la xarxa de proveïment d'aigua de Nogoyá, en tant que aquestes dades queden georeferenciades.

En inglés (máximo 2000 palabras)

Nowadays, especially in developing countries, a large number of medium and small towns are facing the problem of increasing demand for drinking water, although these towns don't have any tools and resources to provide a logical solution from the technical point of view. The use of strategic water supply network models would be a powerful support for the decision-making process for these populations, particularly within the scope of the planning of this infrastructure in the medium and long term. However, the context in which this need usually appears is characterized by the lack of basic information for the application of conventional modeling techniques (among other equally important problems), combined with a bad socio-political and economic situation, which does not allow to reduce this gap in the short term through the usual technical resources. Consequently, a question arises: how to get a mathematical model of a water supply network under these circumstances, which gives a useful outcome for the decision making process in the strategic planning that affects this infrastructure?

This project's main target is to answer this question. To do this, it aims to supplement conventional modeling methodology with the use of secondary information coming from towns different to the town problem to solve, essentially through sorting/classification techniques such as Clustering Analysis and Support Vector Machines. In the calibration phase, it pretends to absorb the uncertainty of a reality that is far away from being described clearly and unequivocally, with the assistance of techniques and concepts of Soft Computing: genetic algorithms and pertinence functions of fuzzy subsets.

This project came out from a study case of the municipality of Nogoyá (Argentina), chosen following a collaboration with the Centro de Estudios para la Gestión Local Sostenible del Agua y del Hábitat Humano, in the Universidad Autónoma de Entre Ríos. A grant from the Centro de Cooperación al Desarrollo de la Universitat Politècnica de València made it possible.

Nogoyá presents some peculiarities that require the achievement of a number of partial objectives. They are specially limited to collect all the available information (this objective was met almost entirely during the 3 months stay in Parana (Entre Rios)) and process it in order to keep it in a functional format. To do this, a project is started on a Geographic Information System, with the aim of creating a great potential repository with available data of the Nogoyá water supply network, since these data are geo-referenced.

Palabras clave (máximo 5): Sistema de distribución de agua, Modelación, Calibración, Datos limitados, Algoritmos genéticos

Este Trabajo Fin de Máster se lo agradezco de corazón...

Al Centro de Cooperación al Desarrollo de la Universitat Politècnica de València, por financiar a través del Programa Cooperación 2014 mi estancia en Paraná (Argentina).

A las instituciones públicas argentinas que han facilitado información o recursos para la elaboración de este estudio. Entre éstas, destacar a la Dirección de Estadística y Censos de la provincia de Entre Ríos por su diligencia; a la Municipalidad de María Grande y en particular a German L. Elberg por su generosidad; y a la Secretaría de Energía de la provincia de Entre Ríos, que hizo posible desplazarme hasta Nogoyá.

Al Centro para la Gestión Local Sostenible del Agua y del Hábitat Humano (CEGELAH). En primer lugar, a su director MSc. Arq. Mario Nudelman, quien se interesó para que saliese adelante la beca del Programa Cooperación 2014, me apoyó en todo lo que pudo durante mi estancia en Paraná y me invitó a su casa, gracias a lo cual conocí a su encantadora familia; a Alejandra Vince y Alejandro Olmos por su amabilidad y ayuda recibida nada más aterrice allí; al Ing. Luis Porcaro, gracias al cual se cerró con satisfacción la instalación del caudalímetro intrusivo y me llevó de vuelta a casa después; y al Ing. Pedro Arriondo, por compartir conmigo sus conocimientos sobre SIG.

A la Municipalidad de Nogoyá y sobre todo al Ing. Daniel Rocca, por su hospitalidad y porque fue quien me proporcionó la mayor parte de la información que ha alimentado este trabajo.

A Bruno Brentan por la ayuda que me prestó en el entrenamiento del SVM y a Enrique Campbell por la información y explicaciones recibidas sobre Análisis Clústering con R y Toolkit de Epanet, ambos colaboradores del grupo de investigación Fluing de la Universitat Politècnica de València. Sin su aportación, el camino andando hubiera sido mucho más tortuoso y lento de transitar.

A Joaquín Sebastián Izquierdo, mi cotutor, por revisar presto el documento de la memoria y por animarme cuando más lo necesitaba, y en especial a Rafael Pérez García, mi tutor, que fue quien abrió la primera puerta para poder llevar a cabo este proyecto, que me ha dado consejos desinteresadamente, de los que luego se recuerdan a lo largo de la vida, y cuyas respuestas a mis dudas me han permitido dar pasos hacia delante con más firmeza.

A Carlos, el hermano de Néstor, por echarme una mano con las deficientes fotos del viaje e interesarse en mi trabajo y, a Cristina, su mujer, quien sabe de Estadística e intentó ayudarme en todo lo que pudo sobre este tema, para lo cual tuvo que escuchar mis divagaciones. A Cristina, la madre de Néstor, por entenderme y compartir conmigo los desayunos a la orilla del mar.

A Sil y Edu. Sin ellos mi estancia en Paraná habría sido triste y solitaria. Me mostraron su tierra: Diamante, la doma, los gauchos, remar por el río Paraná y pasear por su orilla, el dulce de leche... Me hicieron sentir parte de sus vidas presentándome a sus estupendos hijos y nietecito e invitándome a comidas familiares. Les admiro porque hablan de lo que hacen. Son para mí unos amigos muy queridos.

A mi padre, que respeta mis decisiones y me apoya sin preguntas. Espero que considere este pequeño fruto como suyo también, puesto que siempre sembró por y para sus hijos. A mi hermano, que consigue hacerme reír cuando los demás lo tienen muy difícil, le quita cuando puede mucho hierro al asunto y me da esas recomendaciones que tanta falta me hacen. También a su novia, Andrea, que nunca se olvida de preguntar y darme ánimos. A mi madre. Ella ha padecido por mí, conmigo y sin mí lo indecible, si bien, permanece en pie para ser el faro que guía a una hija necesitada de consejos. Y los suyos son impagables por sabios, pues no he conocido a otra persona que haya igualado su sentido común.

A Néstor. Literalmente, él me ha seguido en este maravilloso viaje desde que despegó el avión hacia Argentina hasta la última noche en vela antes de la entrega. Este proyecto sin él habría acabado siendo otra cosa, peor, por descontado. Ha interpretado infinidad de papeles en esta historia: ayudante rozando la esclavitud, tutor, amante, psicólogo, editor, "pensadero", compañero de vida... Ha sufrido también por mí y conmigo, incluso se ha perdido su vigésimo octavo verano... Pese a todo me sigues amando... Cómo no, este TFM antes que a nadie te lo agradezco a ti. ¿Lo sabes, verdad? No lo olvides. Yo no lo olvido.

Índice

1.	Introducción	1
2.	Metodología y herramientas.....	5
2.1.	Técnicas de clasificación	9
2.1.1.	<i>Indicador de disimilaridad/similaridad entre individuos</i>	<i>10</i>
2.1.2.	<i>Algoritmos de clasificación.....</i>	<i>10</i>
2.1.3.	<i>Indicador de calidad del resultado obtenido: La silueta.....</i>	<i>12</i>
2.2.	Proceso de calibración	13
3.	Caso de estudio de Nogoyá.....	15
3.1.	Información general	15
3.1.1.	<i>Ubicación geográfica.....</i>	<i>15</i>
3.1.2.	<i>Caracterización del medio físico natural.....</i>	<i>15</i>
3.1.3.	<i>Caracterización socioeconómica</i>	<i>17</i>
3.1.4.	<i>Análisis demográfico</i>	<i>17</i>
3.2.	Descripción general del sistema de abastecimiento de agua	18
3.3.	Recopilación de información sobre la red de abastecimiento	23
3.3.1.	<i>Información que interviene en la elaboración del modelo.....</i>	<i>23</i>
3.3.2.	<i>Información que interviene en la fase de calibración.....</i>	<i>28</i>
3.3.2.1.	Caudales.....	28
3.3.2.2.	Presiones.....	29
3.4.	Post-proceso de información.....	31
3.4.1.	<i>Tratamiento de información que interviene en la fase de modelación.....</i>	<i>31</i>

3.4.1.1.	Sistema de Información Geográfica (SIG) para el sistema de abastecimiento de agua de Nogoyá	31
3.4.1.2.	Tratamiento de la información que concierne a la estimación del consumo de Nogoyá.....	48
3.4.2.	<i>Tratamiento de información que interviene en la fase de calibración</i>	64
3.4.2.1.	Caudales.....	64
3.4.2.2.	Presiones.....	66
3.5.	Modelación del sistema de abastecimiento.....	67
3.5.1.	<i>Simplificación de la red y empleo de GHydraulics para crear el INP de EPANET</i> 68	
3.5.2.	<i>Puntos de producción</i>	69
3.5.3.	<i>Asignación de cargas</i>	70
3.5.3.1.	Consumo contabilizado	70
3.5.3.2.	Consumo no contabilizado	80
3.5.4.	<i>Asignación de mediciones de presión a nudos del esquema simplificado</i>	80
3.5.5.	<i>Clústering de tuberías y nudos</i>	81
3.6.	Calibración	85
3.6.1.	<i>Datos de entrada del algoritmo</i>	85
3.6.1.1.	Modelo de la red en EPANET	85
3.6.1.2.	Datos de calibración.....	87
3.6.2.	<i>Variables de decisión</i>	89
3.6.2.1.	Valores extremos de los distintos tipos de variables de decisión	89
3.6.3.	<i>Función objetivo</i>	91
3.6.3.1.	Términos asociados con observaciones de caudal inyectado y presiones..	92

3.6.3.2. Términos asociados con la demanda base de consumo	94
3.6.3.3. Restricción dura.....	95
3.6.4. <i>Pruebas realizadas</i>	95
3.7. Análisis de resultados	97
4. Conclusiones y futuras líneas de investigación	111
Referencias	115
ANEJOS.....	117
Anejo 1: Clústering de María Grande	117
Anejo 2: Registros de presiones observadas distribuidas por la red de Nogoyá	126
Anejo 3: Clústering de tuberías y nudos.....	132
Anejo 4: Calibraciones realizadas en Nogoyá	137

Índice de figuras

Figura 1: Esquema propuesto del proceso de calibración con Evolver para una red de abastecimiento de agua.	14
Figura 2: Ubicación geográfica de Nogoyá a escala estatal, provincial y departamental.	15
Figura 3: Ubicación de viales de comunicación y arroyos con respecto a la planta urbana de Nogoyá.	16
Figura 4: Número de habitantes totales en Nogoyá por radio censal según Censo 2010.	18
Figura 5: Esquema general de la red de abastecimiento del Municipio de Nogoyá.	20
Figura 6: Edad, material y diámetro de la red y elevación del terreno en la zona de la planta urbana.	22
Figura 7: Esquema del caudalímetro intrusivo y acotación de la distancia H que marca la profundidad de inserción para que el eje del rotor quede a una profundidad de 1/3 del radio. Fuente: Adaptado del Catálogo de ODIN.	28
Figura 8: Instantánea tomada durante la instalación del caudalímetro intrusivo. Fuente: propia.	29
Figura 9: Ubicación en la planta urbana de los puntos donde se han llevado a cabo mediciones de presión. En la tabla de la izquierda aparece la dirección del domicilio facilitada a la que corresponde cada uno de dichos puntos.	30
Figura 10: Conjunto de parámetros que definen un datum. Fuente: Adaptado de SIGMUR, 2005.	32
Figura 11: Diagrama de flujo para la obtención de la capa 105_104_708_SpatialJoinEditTableText_Manz.	35
Figura 12: Ejemplo de punto conflictivo.	36
Figura 13: Diagrama de flujo para la obtención de las capas 512_510_GHydr_TubORI y 513_511_GHydr_NudORI.	38
Figura 14: Diagrama de flujo para la obtención de 617_616_TIN y 618_617_mde.	43

Figura 15: Gráficos de dispersión que enfrentan las variables implicadas en el clústering de parcelas con su correspondiente consumo medio en los meses de noviembre y diciembre, CI incluidos.....	55
Figura 16: Gráficos 3D que muestran las soluciones del clústering con el algoritmo PAM y variables de entrada superficie total de la parcela, superficie construida y habitantes por vivienda.....	57
Figura 17: Gráficos asociados a cada tipo de clústering empleado que contraponen el valor de silueta media del conjunto de clústeres con el número formado de estos para cada métrica utilizada, en su caso.....	58
Figura 18: Solución adoptada para el clústering de parcelas de María Grande.....	60
Figura 19: Captura de pantalla de la web de GEOMAT en la que puede observarse la selección de parcelas.	62
Figura 20: Extracto de uno de los PDF descargados del visor SIG habilitado por GEOMAT..	62
Figura 21: Clasificación de las parcelas de Nogoyá proporcionada por el SVM entrenado y validado.	63
Figura 22: Gráfico donde aparece el caudal inyectado a red la semana 1 (del 2 al 7 de diciembre de 2014) y 2 (del 8 al 12 de diciembre de 2014) de mediciones.	65
Figura 23: Gráfico en el que aparece la curva de caudal inyectado empleada como dato de calibración, resultado de aplicar la media móvil simple con un dato previo a la media de las mediciones tomadas en las semana 1 y 2.	65
Figura 24: Gráfico en el que aparece la curva que describe la cota de energía aguas abajo de la VRP.....	66
Figura 25: A la izquierda está el esquema completo de la red y a la derecha el simplificado, con el ramal incluido recuadrado.	68
Figura 26: Esquema del modelo introducido en EPANET con los puntos de producción señalados.	69
Figura 27: Identificación de los nudos 305201 y 328001 a través de los cuales se quiere tener en cuenta el consumo de las parcelas ubicadas fuera de la planta urbana consolidada, que quedarían comprendidas en los radios censales 1 y 5, pertenecientes a las fracciones 1 y 3, respectivamente.	73

Figura 28: Representación del CV por clase de parcela y prueba de clústering realizada, para un 10 clústeres.....	75
Figura 29: Representación del CV por clase de parcela y prueba de clústering realizada, para un 20 clústeres.....	75
Figura 30: Representación del CV por clase de parcela y prueba de clústering realizada, para un 30 clústeres.....	76
Figura 31: Resultado final del clústering de nudos.....	78
Figura 32: Funciones de densidad correspondientes al consumo por clase definida en el clústering de nudos de consumo.....	79
Figura 33: Ubicación final en el modelo de los registros disponibles de presión medida.....	80
Figura 34: Resultado final del clústering de tuberías (abajo) y nudos (arriba).....	84
Figura 35: Esquema del proceso de calibración con Evolver de la red de suministro de agua de Nogoyá.....	86
Figura 36: Ejemplo de cómo buscar el valor adecuado en la hoja de control a partir del identificador del elemento.....	87
Figura 37: Ejemplo de los valores medios y umbrales admisibles para el caudal inyectado y las presiones, necesarios para el cálculo de la función objetivo.....	88
Figura 38: Cálculo de la penalización PP_{i-h} asociada a las presiones, evaluada para cada hora.....	92
Figura 39: Cálculo de la penalización PQ_{iny-h} asociada a los caudales inyectados, evaluada para cada hora.....	93
Figura 40: Resultados de las calibraciones consideradas como admisibles en referencia al caudal inyectado, consumido y fugado.....	99
Figura 41: Presiones simuladas en los puntos de la red allí donde se habían llevado a cabo observaciones de la misma.....	102
Figura 42: Las tuberías resaltadas en naranja corresponden a aquellos grupos con un rango de variación en la <i>rugosidad</i> más estrecho con respecto al resto. En la figura de la izquierda, los puntos azules tienen un tamaño proporcional al número de parcelas totales que abastecen cada nudo (los más pequeños corresponden a cero, es decir, únicamente son nudos de conexión). En la figura del centro, los puntos morados son los asociados a	

aquellos grupos con coeficientes de descarga cuya media es más alta en relación a la del resto. Por último, en la figura de la izquierda se destacan las tuberías cuya mediana está por encima de 0.2, siendo las de color **fucsia** las que corresponden a diámetros superiores a 100 mm.....103

Figura 43: Resultados obtenidos para las variables de decisión rugosidad y coeficiente de descarga..... 104

Figura 44: Los puntos **morados** son los asociados a aquellos grupos con coeficientes de descarga cuya media es más alta en relación a la del resto. Los puntos **verdes** tienen un tamaño proporcional a la longitud ponderada de tubería y los **azules** al número de acometidas, ambos valores estimados para el clústering de coeficientes de descarga. La zona con un número de acometidas y longitud ponderada de tubería en media alto está rodeada por una circunferencia. En la figura de la izquierda se resaltan los nudos en **naranja** que forman parte de clústeres con intervalos intercuartílicos menores a 0.05.....105

Figura 45: Resultados obtenidos para la variable de decisión *factores de demanda*.....105

Figura 46: Resultados obtenidos para la variable de decisión *demanda base* de los nudos de consumo.....106

Figura 47: Resultados de presiones mínimas y máximas obtenidas con las distintas configuraciones encontradas del modelo.....108

Figura 48: Se encuentran resaltados, en la figura de la izquierda, los puntos en los que se dan presiones mínimas por debajo de 5 mca y, en la de la derecha, los de presión máxima por encima de 20 mca.....108

Figura 49: Gráficos de resultados correspondientes a los puntos que disponen de presiones observadas y figura en la que puede consultarse su ubicación, descartados por distintas razones.....109

Figura A1: Registro de presiones para la época primavera-verano asociado a la medición P1, ubicada en el domicilio con dirección 17 de Agosto 1967. El medidor se instaló en una acometida conectada directamente a una tubería de distribución de diámetro 63 mm.....126

Figura A2: Registro de presiones para la época primavera-verano asociado a la medición P2, ubicada en el domicilio con dirección Avda. Marchini y Bv. Alberdi. El medidor se instaló en una acometida conectada directamente a una tubería de distribución de diámetro 160 mm.126

- Figura A3:** Registro de presiones para la época primavera-verano asociado a la medición P3, ubicada en el domicilio con dirección Bv. España 1346. El medidor se instaló en una acometida conectada directamente a una tubería de distribución de diámetro 125 mm.127
- Figura A4:** Registro de presiones para la época primavera-verano asociado a la medición P4, ubicada en el domicilio con dirección Bv. España y 25 de Mayo. El medidor se instaló en una acometida conectada directamente a una tubería de distribución de diámetro 200 mm. ...127
- Figura A5:** Registro de presiones para la época primavera-verano asociado a la medición P5, ubicada en el domicilio con dirección F. Gerald y Q. y Taboada 876. El medidor se instaló en una acometida conectada directamente a una tubería de distribución de diámetro 160 mm.128
- Figura A6:** Registro de presiones para la época primavera-verano asociado a la medición P6, ubicada en el domicilio con dirección J.J. Paso y Marchini. El medidor se instaló en una acometida conectada directamente a una tubería de distribución de diámetro 63 mm.....128
- Figura A7:** Registro de presiones para la época primavera-verano asociado a la medición P7, ubicada en el domicilio con dirección Avda. Marchini y Colon. El medidor se instaló en una acometida conectada directamente a una tubería de distribución de diámetro 200 mm.129
- Figura A8:** Registro de presiones para la época primavera-verano asociado a la medición P8, ubicada en el domicilio con dirección Avda. Marchini 1582. El medidor se instaló en una acometida conectada directamente a una tubería de distribución de diámetro 200 mm. ...129
- Figura A9:** Registro de presiones para la época primavera-verano asociado a la medición P9, ubicada en el domicilio con dirección Q. y Taboada 876. El medidor se instaló en una acometida conectada directamente a una tubería de distribución de diámetro 200 mm. ...130
- Figura A10:** Registro de presiones para la época primavera-verano asociado a la medición P10, ubicada en el domicilio con dirección Ruta 12 y Victoria. El medidor se instaló en una acometida conectada directamente a una tubería de distribución de diámetro 63 mm.....130
- Figura A11:** Registro de presiones para la época primavera-verano asociado a la medición P11, ubicada en el domicilio con dirección Yrigoyene/Parana y S. Oriz. El medidor se instaló en una acometida conectada directamente a una tubería de distribución de diámetro 175 mm..... 131
- Figura A12:** Siluetas y gráficos de puntos asociados al clústering de nudos, a través de los cuales se pueden identificar los distintos grupos conformados según tipo de clústering y combinaciones de pesos (aplicados a las variables de entrada) para el número máximo de grupos considerado, 30..... 134

Figura A13: Siluetas y grafo de la red asociados al clústering de tuberías, a través de los cuales se pueden identificar los distintos grupos conformados según tipo de clústering y combinaciones de pesos (aplicados a las variables de entrada) para el número máximo de grupos considerado, 25.....136

Índice de tablas

Tabla 1: Distancias consideradas como indicador de disimilaridad.....	10
Tabla 2: Métodos que definen la forma de evaluar la distancia inter-clúster.	11
Tabla 3: Descripción de los algoritmos KMEANS y PAM.....	12
Tabla 4: Características y dimensiones de los tanques 35 y 33.	21
Tabla 5: Características técnicas de las electrobombas sumergibles instaladas en los distintos pozos (datos de octubre de 2014).	21
Tabla 6: Listado de información con prioridad 1 acerca de la red de distribución de Nogoyá.	24
Tabla 7: Listado completo de información requerida a la contraparte para la elaboración del modelo. Descripción de prioridad, disponibilidad, formato y fuente principal de información para cada ítem.	27
Tabla 8: Definición del Sistema de Coordenadas Planimétrico POSGAR 94 y proyección Gauss-Krüger para la Faja 5.	33
Tabla 9: Resultados obtenidos en el Bloque 1 del SIG, Trama Urbana.	34
Tabla 10: Información de partida para el desarrollo del Bloque 1 del SIG, Trama Urbana.....	34
Tabla 11: Resultados obtenidos en el Bloque 2 del SIG, Topología de la Red.....	37
Tabla 12: Información de partida para el desarrollo del Bloque 2 del SIG, Topología de la Red.....	38
Tabla 13: Resultados obtenidos en el Bloque 3 del SIG, Elevación del Terreno.....	42
Tabla 14: Información de partida para el desarrollo del Bloque 3 del SIG, Elevación del Terreno.	42

Tabla 15: Resultado obtenido en el Bloque 4 del SIG, Censo 2010.....	47
Tabla 16: Información de partida del Bloque 4, Censo 2010.	47
Tabla 17: Información proporcionada por la Municipalidad de María Grande.....	49
Tabla 18: Archivos consultados del Censo 2010 para la determinación del ratio habitantes por vivienda y radio censal.	49
Tabla 19: Resultados alcanzados al introducir en ArcGIS la información de DMG02, DMG05, DC02 y DC03.....	51
Tabla 20: Resumen del análisis univariado de la <i>superficie total de la parcela</i> (Sup_m2).	52
Tabla 21: Resumen del análisis univariado de la <i>superficie construida</i> (Edif_m2).	53
Tabla 22: Resumen del análisis univariado de los <i>habitantes por vivienda</i> (HabxViv).	53
Tabla 23: Resumen del análisis univariado del <i>consumo medio en el bimestre noviembre-diciembre</i> (Cnd_13_m3xmes).....	54
Tabla 24: Coeficiente de Correlación de Pearson entre el consumo y las variables consideradas para el clústering de parcelas.....	55
Tabla 25: Resumen de la distribución temporal de las mediciones de presión disponibles. Las casillas coloreadas de azul corresponden a la época otoño-inverno y de naranja a la época primavera-verano.....	67
Tabla 26: Resultados obtenidos en el Bloque 5 del SIG, Demanda.....	71
Tabla 27: Tipos de clústering, combinación de pesos y número de clústeres combinados en las pruebas llevadas a cabo de clústering de tuberías y parcelas.	83
Tabla 28: Valores de los umbrales para el caudal inyectado y las presiones observadas.....	88
Tabla 29: Número total de elementos por cada tipo de variable de decisión.....	89
Tabla 30: Rangos mínimos y máximos para las variables de decisión (excepto la <i>demand base</i>).	90
Tabla 31: Valores mínimos y máximos para las demandas base de nudos de consumo, diferenciados por clúster.	91
Tabla 32: Valores posibles de los parámetros del algoritmo genético.	96

Tabla 33: Valores establecidos para los criterios de parada del algoritmo genético.	96
Tabla 34: Calibraciones realizadas para la obtención de diferentes configuraciones admisibles del sistema de suministro de agua de Nogoyá.....	97
Tabla 35: Para las calibraciones lanzadas que obtienen resultados admisibles, valores alcanzados por la función objetivo y por cada uno de los término que forman parte de ésta, siendo P=Progreso y NME=Número máximo de evaluaciones.	98
Tabla A1: Gráficos de dispersión Edif_m2 vs. Sup_m2 y de la silueta correspondiente al clústering jerárquico aglomerativo, método ward.D2, para 5 clústeres y cada métrica probada.	118
Tabla A2: Gráficos de dispersión Edif_m2 vs. Sup_m2 y de la silueta correspondiente al clústering jerárquico aglomerativo, método UPGMA, para 5 clústeres y cada métrica probada.	120
Tabla A3: Gráficos de dispersión Edif_m2 vs. Sup_m2 y de la silueta correspondiente al clústering jerárquico divisivo DIANA, para 5 clústeres y cada métrica probada.	122
Tabla A4: Gráfico de dispersión Edif_m2 vs. Sup_m2 y de la silueta correspondiente al clústering de partición KMEANS, para 5 clústeres.	123
Tabla A5: Gráficos de dispersión Edif_m2 vs. Sup_m2 y de la silueta correspondiente al clústering de partición PAM, para 5 clústeres y cada métrica probada.	125
Tabla A6: Valores de la función objetivo y cada término de ésta para todas las calibraciones lanzadas, siendo P=Progreso y NME=Número máximo de evaluaciones.....	137

1. Introducción

La infraestructura de abastecimiento de agua potable juega un papel fundamental en la vida de los seres humanos a los que da servicio. El gestor de dicha infraestructura debe conocer el estado y planificar la evolución de ésta para asegurar su sostenibilidad en el tiempo, lo cual significa que de forma continuada se proporciona agua a los habitantes de una población en cantidad suficiente, con la calidad necesaria y a un coste razonable, tanto desde el punto de vista económico como medioambiental.

Los modelos estratégicos de la red de suministro de agua son un instrumento básico que asiste a la planificación a medio y largo plazo de la infraestructura de abastecimiento de agua. Constituyen el fundamento de los estudios que requieren dicha planificación, sin cuyos resultados las decisiones tomadas carecen de un soporte técnico robusto. No obstante, la elaboración de esta herramienta solicita una cantidad de información notablemente exhaustiva en el espacio y el tiempo, que no siempre está disponible.

En la actualidad, sobre todo en países en vías de desarrollo, un gran número de medianas y pequeñas poblaciones se enfrentan al aumento de la demanda o al envejecimiento de su infraestructura de suministro de agua, si bien no poseen un instrumento como los modelos matemáticos de la red para dar solución a estos problemas con una base sólida según criterios técnicos. Esta necesidad se plantea en un contexto caracterizado por la escasez, entre otros no menos importantes, de información básica para la aplicación de técnicas de modelación convencionales, sumada a una situación sociopolítica y económica que no permite paliar esta carencia en el corto plazo a través de los recursos técnicos habituales.

El presente Trabajo Fin de Máster (TFM) gira entorno a un caso de estudio que queda enmarcado por la realidad esbozada en el anterior párrafo. El nombre de la población a la que hace referencia es Nogoyá, un municipio argentino perteneciente a la Provincia de Entre Ríos. Su servicio de abastecimiento suministra agua de forma continua, si bien la información referente a la red es en más de un aspecto deficitaria: El catastro no está actualizado ni en soporte digital fácilmente accesible ni, en su defecto, funcional; existe información de importancia (por ejemplo, los partes de roturas) únicamente en soporte papel; carece de micromedición y de estudios sobre consumo fundamentados en mediciones que permitan hacer balances hídricos de cierta precisión; no se monitorean las presiones de servicio en red y no dispone de dispositivos de medición del caudal inyectado. La Municipalidad de Nogoyá es consciente de las carencias anteriormente expuestas y de que deben ser subsanadas para afrontar los nuevos retos que plantean los presentes y futuros desarrollos urbanísticos. No obstante y en relación a esto último, pese a la Ley

Provincial 10.027¹, Nogoyá no dispone de un plan de ordenamiento urbano que regule de forma eficaz el uso y ocupación del suelo y que, en consecuencia, fuerce a una planificación estratégica que asegure la sostenibilidad del servicio de agua potable. Existen esfuerzos por avanzar en la normalización del desarrollo urbanístico (tal y como demuestra la Ordenanza N° 962)² y en estudios que guíen la toma de decisiones en materia de gestión a medio y largo plazo de las infraestructuras municipales, pero evoluciona de forma más rápida la construcción de nuevas viviendas³.

Sobre este escenario, surge una colaboración entre el Centro de Investigación para la Gestión Local Sostenible del Agua y el Hábitat Humano (CEGELAH), adscrito a la Facultad de Ciencias y Tecnología de la Universidad Autónoma de Entre Ríos (UADER), y la Municipalidad de Nogoyá, que pretende dar respuesta a las inquietudes planteadas sobre la gestión de la infraestructura de abastecimiento de agua potable. Ésta se concreta en el proyecto *"Desarrollo, Asesoramiento y Capacitación no formal de un Modelo de Simulación de la Red de Abastecimiento de Agua del Municipio de Nogoyá"* (al que se hace referencia en este TFM con las siglas DAMRAN), planteado al Área de Gestión y Transferencia Tecnológica de la UADER como un servicio a terceros. En él se prevé participen como grupo de trabajo el Ing. Luis Porcaro, el Ing. Daniel Rocca, el Ing. Pedro Arriondo y el MSc. Arq. Mario Nudelman. Los objetivos generales que persigue son:

- a) *"Implantar en el servicio de distribución de agua y saneamiento del Municipio de Nogoyá un sistema de información local, que permita la elaboración de Diagnósticos y Planes de Mejora Continua orientados a la consecución de objetivos de Gestión Sostenible del Agua y Eficiencia Hidro-Energética (EH)."*
- b) *"Capacitar directa o indirectamente los recursos humanos locales para la operación y mantenimiento del sistema, que incorpora los programas informáticos e instrumental"*

¹ Afecta al régimen municipal y especifica en su artículo 7º que son los municipios los agentes que deben "dicta normas de ordenamiento territorial" dentro de los límites del ejido, en las que se distingan "zonas urbanas, suburbanas y rurales", estableciéndose a su vez en cada una de éstas "zonas de subdivisión, usos e intensidad de ocupación del suelo, en pos del desarrollo local sostenible y la mejora de la calidad de vida de su población".

² La Ordenanza N° 962, de septiembre de 2012, dicta en su artículo 17º, apartado b, que en las zonas de nueva urbanización sea factible y posible la conexión a los servicios de agua potable.

³ Esto es en gran parte debido a programas que facilitan el crédito, como PRO.CRE.AR (Ordenanza N° 1.030) o PRO.MU.VI.SO (Ordenanza N° 1.067), los cuales indiscutiblemente resuelven el problema, también de vital importancia, que sufren aquellos ciudadanos que no pueden acceder por otras vías a una vivienda de calidad habitacional digna.

asociado orientados al logro de una Gestión Sostenible del servicio a la EH del mismo."

- c) *"Introducir a los Decisores Políticos, Técnicos y Personal de planta implicado a nociones y procesos de mejora continua y auto-regulación del sistema."*

Cabe añadir que la relación entre el CEGELAH y la Municipalidad de Nogoyá tiene un antecedente previo: la elaboración del informe técnico con título *"Diagnóstico sobre el uso de energía eléctrica en los abastecimientos de agua y saneamiento. Experiencia piloto en el Municipio de Nogoyá."*, fruto de un convenio entre la Secretaria de Energía de la Provincia de Entre Ríos y la UADER. En la actual investigación se emplearán algunos de los resultados de dicho informe.

En virtud de una beca del Programa Cooperación 2014, otorgada por el Centro de Cooperación al Desarrollo de la Universitat Politècnica de València, fue posible realizar una estancia de tres meses en la ciudad de Paraná (desde mediados de septiembre de 2014 a mediados de diciembre de 2014) para colaborar con el CEGELAH en el marco del proyecto DAMRAN.

En línea con todo lo anteriormente expuesto, el objetivo principal es conseguir un modelo matemático de una red de abastecimiento de agua que arroje resultados útiles para la toma de decisiones en la planificación estratégica que afecta a esta infraestructura, partiendo de una situación de información disponible *limitada*. Para ello, se pretende complementar la metodología convencional de modelación con el aprovechamiento de información secundaria y procedente de poblaciones que se sitúan en la misma región que la del problema a resolver, gracias a la asistencia de técnicas estadísticas para el análisis de datos y el uso de herramientas englobadas dentro del Soft Computing. No obstante, los procedimientos descritos en este TFM están estrechamente ligados al propósito de proporcionar al municipio de Nogoyá un modelo estratégico de su red. En consecuencia, parten de unas condiciones específicas de información disponible *limitada* y, en este sentido, no pretenden dar respuesta a cualquier casuística que pueda ofrecerse. Si bien, posiblemente provean de ideas para la resolución de problemas de la misma índole o tengan difusión a nivel provincial, ya que es probable que otros municipios pequeños y medianos de Entre Ríos se encuentren en una situación similar a la de Nogoyá.

Por otro lado, el caso de estudio presenta particularidades que requieren de la consecución de una serie de objetivos parciales, que se circunscriben principalmente a recabar toda la información que esté disponible (este objetivo se cumplió casi en su totalidad durante los 3 meses de estancia en Paraná (Entre Ríos)) y procesarla para que quede almacenada en un formato útil.

El documento de la memoria, en el que se desarrolla el camino andado para la consecución de los objetivos mencionados, se estructura en 5 puntos. El primero lo constituye esta

misma introducción, en la que se expone los condicionantes del problema a resolver, el objetivo principal perseguido y la idea vertebradora detrás de la metodología para resolverlo.

En el segundo capítulo se describe con mayor profundidad la metodología propuesta, añadiendo explicaciones en detalle sobre los conceptos que implica allí donde se creen convenientes.

La tercera sección es la más voluminosa y hace referencia al caso de estudio. En ella se describen todos los pasos dados, desde la recopilación de información hasta el análisis de resultados, para la obtención del modelo estratégico de la red de Nogoyá. Es conveniente añadir que en algunos aspectos el contenido está expuesto con la intención de que sirva de guía, de tal manera que quien desee darle continuidad al presente trabajo, modificarlo o mejorarlo, conozca con cierta precisión cómo se han obtenido los resultados y cómo puede reproducir las operaciones realizadas en el caso de que lo necesite.

El último capítulo contiene las conclusiones del estudio y posibles líneas de investigación futuras.

2. Metodología y herramientas

Existe una cantidad no despreciable de información contenida en fuentes que al principio pueden ser descartadas por calificarlas de poco fiables o demasiado imprecisas. En un contexto en el que la información disponible es escasa, se debe sacar rendimiento al anterior hecho. Así pues, estudios realizados por entes gubernamentales (como institutos de estadística o cartografía) a distinta escala respecto a la requerida, registros locales imprescindibles para la recaudación impositiva (por ejemplo, el Catastro municipal), el conocimiento que poseen los operarios de un pequeño-mediano sistema de suministro de agua transmitido en una conversación, incluso Google Maps, deben ser tenidos en cuenta, aunque finalmente sólo sirvan para cotejar una información con otra.

Por descontado, un sistema de abastecimiento de agua potable normalmente es complejo por definición. En consecuencia, la cantidad de información necesaria para poder modelarlo suele ser considerable. En este estudio se parte de:

- Un conocimiento medianamente bueno de la topología de la red y de las características básicas de las tuberías y nudos de conexión (cota). No obstante, la forma en la que está presentada en general toda esta información no es aprovechable directamente.
- Los puntos de producción están definidos, aunque con poca precisión.
- No existe ningún dato robusto acerca del consumo. Si bien, se dispone de datos de micromedición de una población de menor tamaño y de la misma zona que la del problema a resolver, emparejados con la superficie total de la parcela y la superficie construida correspondientes. Es de importancia recalcar, que en el caso de estos municipios, una parcela corresponde habitualmente a una vivienda.
- Sólo se lleva a cabo la medición de una parte del caudal inyectado en un momento muy puntual del año.
- Las mediciones de presión distribuidas por la red se toman de forma espaciada en el tiempo, los registros sólo cubren parte de la semana y su ubicación es imprecisa.

El primer paso que hay que dar cuando información imprescindible está dispersa y en formatos poco útiles es post-procesarla. Para el caso particular de los servicios de abastecimiento de agua, es conveniente recomendar el almacenamiento de datos en un Sistema de Información Geográfica (SIG), ya que en la actualidad ha quedado patente la pujanza que radica en disponer de los mismos referenciados espacialmente (Bartolín, 2013). Los SIG no son objeto del presente estudio. Se reconoce que del mismo no se aprovecha todo su potencial, simplemente se ha utilizado como un instrumento auxiliar. El principal software empleado en el post-proceso de información ha sido el ArcGIS. En el apartado 3.4.1.1 se describe una relación de las operaciones realizadas con dicho software, especificando las herramientas empleadas, y los resultados obtenidos asociados al caso de

estudio. Para la validación de la topología de la red se ha usado el programa GISRed 1.0 y, en tanto que EPANET (Rossman, 2001) es el software adoptado para el análisis del comportamiento de la misma, con la extensión de QGIS GHydraulics se obtiene el archivo de entrada de EPANET (INP) a partir de las capas editadas en ArcGIS.

Una vez se organiza la información y se vuelca a un soporte funcional, se lleva a cabo la simplificación inherente a un modelo, que en algunos casos debe adaptarse a los datos disponibles. No obstante, existen partes de la misma no directamente representable en dicho modelo y que son de importancia *secundaria* en esta etapa, como, por ejemplo, la edad y material de las tuberías. Además, hay que tratar los datos de micromedición procedentes de otra población distinta a la del problema a resolver, para encontrar alguna solución al hecho de que el consumo está pobremente caracterizado. Las técnicas empleadas para convertir este tipo de información en una oportunidad están asociadas a la clasificación de individuos de una muestra, en concreto, al Análisis de Clústering y los Support Vector Machines (SVM), junto a la asistencia de conceptos estadísticos, como la generación aleatoria de muestras y la estimación no paramétrica de la función de densidad, que serán ampliados en el apartado 2.1. Los procedimientos seguidos, que para el caso de estudio se desarrollan en los apartados 3.4.1.2, 3.5.3 y 3.5.5, se especifican a continuación.

Demanda base de los nudos de consumo

La hipótesis de partida que vertebra este procedimiento es que resulta esperable que propiedades inmobiliarias (parcelas) de características similares, en distintas poblaciones de la misma región, posean como grupo una variabilidad parecida en relación al consumo. Así pues, en primer lugar, se busca mediante técnicas de clústering una estructura en los datos de micromedición disponibles en relación a la superficie construida y total de cada parcela/vivienda. La clasificación obtenida sirve como dato de entrada para el entrenamiento y validación de un SVM. Una vez se consigue y prepara la información que caracteriza a las parcelas del caso a resolver, se clasifican con el SVM. Seguidamente, se aplica la técnica de asignación de cargas más conveniente, con el objetivo de adjudicar a cada nudo de consumo las parcelas de cada clase que le corresponden. Dichos nudos de consumo se ubican en la red intentando que estén distribuidos uniformemente a lo largo de la misma y su cantidad viene delimitada por el umbral mínimo de 20 parcelas por nudo de consumo (impuesto en un intento de que una modulación suave del consumo quede justificada). El siguiente paso es aplicar un análisis clústering a estos últimos, con la intención de agruparlos en función de la cantidad de parcelas de cada clase que llevan asociadas. Para cada clase resultado y a partir de los datos de micromedición disponibles, se generan aleatoriamente 1000 individuos de la variable *demanda base de los nudos de consumo de la clase i* , muestra que permite definir una función de densidad que caracteriza dicha variable. Esta última se empleará a su vez como dato de entrada en el proceso de calibración.

Clústering de tuberías y nudos

En un intento de reducir el número de grados de libertad en el proceso de calibración, se ha realizado un clústering de tuberías relacionado con su rugosidad y uno de nudos tocante a los coeficientes de descarga. Ambos se basan en la idea de que elementos de características parecidas sometidos a condiciones similares se comportaran de forma análoga. Las variables empleadas hacen referencia en algunos casos a dicha *información secundaria* que no se ha visto incorporada hasta ahora en el modelo. Éstas son:

1. Clústering de tuberías. Siguiendo las ideas expuestas en Echávez (1997) y propuestas por Mostkov en Idelchik (1994), el diámetro, material y edad de la tubería, así como la calidad del agua y las condiciones hidráulicas a las que está sometida a lo largo de su vida útil, son factores determinantes en la evolución de su rugosidad. Es por ello que se tienen en cuenta las siguientes variables:
 - *Coordenadas del punto medio de la tubería* (2 variables numéricas). La ubicación de este elemento en distintas zonas de la planta urbana, con diferentes grados de consolidación, y en relación a los puntos de producción, condiciona el régimen hidráulico al que está sometido y, por tanto, la velocidad y caudal que discurren a su través.
 - *Material, edad y diámetro* (2 variables nominales y 1 numérica, respectivamente).
2. Clústering de nudos. Si se consulta los distintos términos del Índice Estructural de Fugas (ILI son sus siglas en inglés) en Winarni (2009) queda explicada la elección de las variables:
 - *Coordenadas y cota* (3 variables numéricas), ambas relacionadas indirectamente con la presión a la que está sometido el nudo.
 - *Longitud de tubería ponderada*, considerando el esquema completo de la red, y *densidad de acometidas* (2 variables), estimadas a partir del número de viviendas.

El software utilizado para ejecutar los procedimientos expuestos es R Statistics (en especial para el Análisis Clústering), la librería de Matlab LibSVM (que también está implementada en el paquete de R *e1071*) y hojas de cálculo, en una de las cuales se ha escrito una subrutina para generar la muestra aleatoria de 1000 individuos.

Llegados a este punto, puede observarse que ya se ha hecho referencia a tres etapas de la fase del modelo dentro del proceso de modelación: la recopilación de información, la esqueletización de la red y el análisis y asignación de consumos. Por tanto, faltaría llevar a cabo la última operación definida para los consumos no registrados (cuarta etapa). Estos se reducen en el presente estudio a las fugas, que serán representadas en el modelo por emisores, cuyos coeficientes de descarga constituyen una variable de decisión en el proceso

de calibración. En López *et al.* (1996) se pueden encontrar más detalles sobre cada una de las cuatro etapas nombradas.

En consecuencia, sólo resta la *fase de calibración*. En la misma se hace necesario responder: ¿cómo absorber la incertidumbre de una realidad que dista mucho de estar descrita de forma clara y unívoca a la hora de modelarla? Para ello se acude a técnicas y conceptos propios del Soft Computing: los algoritmos genéticos y las funciones de pertenencia de los subconjuntos difusos.

Los algoritmos genéticos se han empleado con el propósito de obtener varias configuraciones del sistema (posibles modelos diferentes) que obtengan un valor de la función objetivo considerado como admisible. Estas configuraciones y sus respectivos resultados se estudiarán para estimar el rango de variación de las variables que intervienen en ambos. Dichos rangos podrían interpretarse como bandas de confianza. Los algoritmos genéticos son una herramienta potente en tanto que permiten enfrentar problemas no lineales de alta complejidad por la cantidad de variables de decisión combinadas (tal y como el que presenta la calibración de una red), así como llevar a cabo búsquedas múltiples en el espacio de solución, lo cual, junto con el hecho de que utilizan reglas de transición probabilísticas, asegura el poder encontrar distintos sub-espacios de soluciones útiles (Valero, 2010). Resta añadir que en el presente trabajo, tal y como se ha hecho en otros dentro del mismo contexto (Takahashi *et. al.*, 2010), se introduce la demanda base de los nudos de consumo como variable de decisión en el proceso de calibración/búsqueda. A causa de ello, la cantidad de las mismas se incrementa considerablemente, de ahí que sea interesante llevar a cabo un clústering de tuberías y nudos.

El concepto de funciones de pertenencia de los subconjuntos difusos se toma para introducir en la función objetivo la incertidumbre asociada a las observaciones de presiones y de caudal inyectado (datos de calibración), así como la asociada a la demanda base de los nudos de consumo. En consecuencia, no se pretende, por ejemplo, que los resultados de cada individuo probado en el proceso de búsqueda reproduzcan con exactitud los valores observados, ya que los registros de los que proceden se han empleado asumiendo fuertes hipótesis que introducen incertidumbre. Por el contrario, únicamente se intenta discernir entre resultados "aceptablemente correctos", "medianamente correctos" y "no admisibles". Esta forma "vaga" de calificar un resultado, permite absorber parte de dicha incertidumbre. La función de pertenencia define un subconjunto difuso A del referencial X en tanto que empareja un elemento de X con un valor del intervalo [0, 1], el cual calibra en qué grado dicho elemento forma parte de A. Esto es precisamente lo que se pretende al definir funciones de penalización que tienen como input los resultados y cuyo output determina el valor alcanzado por la función

objetivo. Dichas funciones se han definido, en el caso de la demanda base de los nudos de consumo a partir de las funciones de densidad obtenidas anteriormente, mientras que en el caso de las presiones y el caudal inyectado observados a partir de las mediciones disponibles.

El software empleado para implementar estas ideas es Evolver, un programa auxiliar de Microsoft (MS) Excel que tiene como principal propósito la resolución de problemas de optimización mediante algoritmos genéticos, aunque en este caso se persigue proporcionar resultados a un problema de búsqueda. El enlace entre la herramienta de MS y Evolver facilita el empleo del Toolkit de EPANET a través del lanzamiento de distintas macros en etapas concretas del proceso completo de búsqueda. En el apartado 2.2 se desarrollan todas estas ideas con más detalle.

Hay que tener presente que detrás de los conceptos aprovechados en la anterior metodología existen cuerpos teóricos amplios, que ofrecen distintas posibilidades en cada caso. Ninguno de ellos es el objeto principal del presente TFM y, debido a ello, no se ha realizado una revisión exhaustiva de los mismos. Seguidamente, se profundiza únicamente en aquellas técnicas y conceptos que se han empleado en particular para este estudio.

2.1. Técnicas de clasificación

En general, un proceso de clústering consta de los siguientes elementos principales:

1. Partir de un conjunto de N individuos de los que se dispone de una información cifrada por un conjunto de n variables (una matriz de datos de $N \times n$). Éstas deben ser escogidas cuidadosamente, ya que introducir en el proceso alguna no relevante puede generar ruido y falsear los resultados.
2. Establecer un criterio o indicador de disimilaridad o similaridad entre individuos (matriz de $N \times N$). El criterio o índice se aplica o calcula antes del proceso de clasificación. Cabe aclarar que existen algoritmos, como KMEANS, que no emplean este indicador.
3. Escogemos un algoritmo de clasificación para determinar la estructura de agrupación de los individuos.
4. Establecer un criterio o indicador de similaridad entre clústeres. Este elemento se encuentra sólo en los algoritmos de clasificación jerárquicos aglomerativos implementados en R. El criterio o índice se aplica o calcula durante el proceso de clasificación.

2.1.1. Indicador de disimilaridad/similaridad entre individuos

El concepto de distancia es el que se ha tomado como medida de disimilaridad/similaridad entre individuos para este estudio. Los tipos de distancias tenidas en cuenta en aquellos algoritmos de clasificación que lo permitían se pueden consultar en la **Tabla 1**.

	Fórmula	Aclaración
Distancia Euclidiana	$d_2(x_i, x_j) = \ x_i - x_j\ _2 = \sqrt{(x_i - x_j)'(x_i - x_j)} = \sqrt{\sum_{l=1}^n (x_{il} - x_{jl})^2}$	Función <i>daisy</i> del paquete <i>cluster</i> .
Distancia Manhattan	$d_1(x_i, x_j) = \sum_{l=1}^n x_{il} - x_{jl} $	Función <i>daisy</i> del paquete <i>cluster</i> .
Distancia de Mahalanobis	$\ x_i\ _B = \sqrt{x_i' B x_i}$	Función <i>D2.dist</i> del paquete <i>biotools</i> . B es la matriz de varianzas-covarianzas de las variables.
Distancia Gower	$d_{ij}^2 = 1 - s_{ij}$ $s_{ij} = \frac{\sum_{h=1}^{p_1} (1 - x_{ih} - x_{jh} /G_h) + a + \alpha}{p_1 + (p_2 - d) + p_3}$	Función <i>daisy</i> del paquete <i>cluster</i> . p ₁ es el número de variables cuantitativas continuas, p ₂ es el número de variables binarias, p ₃ es el número de variables nominales, a es el número de coincidencias (1,1) en las variables binarias, d es el número de coincidencias (0,0) en las variables binarias, α es el número de coincidencias den las variables nominales y Gh es el rango de la h-ésima variable cuantitativa.

Tabla 1: Distancias consideradas como indicador de disimilaridad.

Cabe destacar que las distancias Euclidiana y Mahattan se ven afectadas por el orden de magnitud de la variable. Una forma de soslayar este problema es tipificar las mismas previamente al cálculo de la matriz de disimilaridades, si bien dicha operación no está recomendada en cualquier caso y hay que evaluar su idoneidad. La distancia de Mahalanobis, por contra, es invariante frente a cambios de escala y la distancia Gower lleva a cabo una normalización de las variables numéricas, al dividir entre el rango que toman dentro de la muestra analizada. Esta última, además, permite incluir en el análisis variables nominales y binarias.

2.1.2. Algoritmos de clasificación

Los algoritmos de clasificación empleados en este TFM se agrupan seguidamente.

Clustering jerárquico

Esta técnica construye un árbol, llamado *dendograma*, que representa a qué medida de disimilaridad (longitud de la *clado*) se van uniendo entre sí, en el primer nivel empezando desde abajo, individuos y, a partir del segundo, grupos. La exploración de todos los

posibles árboles es computacionalmente intratable, por lo tanto, suelen seguirse algoritmos aproximados guiados por determinadas heurísticas. Existen dos aproximaciones diferentes al clustering jerárquico:

- *Clustering jerárquico aglomerativo*: se comienza con tantos clústeres como individuos y consiste en ir formando (aglomerando) grupos según la similitud entre ellos. Las funciones implementadas en R *hclust* y *agnes* (paquete *cluster*) llevan a cabo este tipo de clustering.
- *Clustering jerárquico divisivo*: se comienza con un único clúster y consiste en ir dividiendo clústeres según la disimilitud entre sus componentes. La función empleada para aplicar este clúster es *diana* (paquete *cluster*).

En el anterior apartado, ya se ha señalado de qué forma se calibra la disimilaridad entre individuos. La pregunta que surge ahora es cómo evaluar la similaridad entre agrupaciones en el caso del clustering jerárquico aglomerativo. Existen varios métodos que dan respuesta, basados en distintos modos de evaluar la distancia inter-clúster. En la **Tabla 2** se describen los utilizados en este estudio.

Método	Descripción
Ward.D2	<p>En cada paso del análisis, considera la posibilidad de la unión de cada par de grupos y optar por la fusión de aquellos dos que menos incrementen la siguiente función de optimización (Murtagh, 2014):</p> $D(c_1, c_2) = \delta^2(c_1, c_2) = \frac{ c_1 c_2 }{ c_1 + c_2 } \ c_1 - c_2\ ^2$ <p>Donde c_1 y c_2 hacen referencia a cada clúster y la fórmula mide el cambio en la suma total de los cuadrados de las distancias a causa de la fusión.</p>
UPGMA ("average")	<p>Utiliza como distancia entre un grupo l y un individuo j la media de las distancias entre los individuos del grupo l y el individuo j. La definición se extiende a la distancia entre dos grupos como la media de todas las distancias entre todos los pares de individuos de los dos grupos. Este método es espacio-conservativo, es decir, no hace variar considerablemente la métrica inicial.</p>

Tabla 2: Métodos que definen la forma de evaluar la distancia inter-clúster.

Clustering de partición

Éstas técnicas realizan una distribución de los elementos entre un número prefijado de clústeres o grupos. En este trabajo se han empleado 2 algoritmos KMEANS y PAM (paquete *cluster*), cuyas etapas se definen en la **Tabla 3**. La diferencia principal entre ambos es que Kmeans no trabaja con las disimilaridades entre individuos directamente, ya que su objetivo es minimizar la suma del cuadrado de la distancia Euclidiana entre individuos asignados a un clúster y el centroide del mismo. PAM, en cambio, sí usa las disimilaridades entre individuos para determinar los llamados *medoids* (que son

individuos de la propia muestra) y de ahí su potencia, pues no está acotado a la distancia Euclidiana y, por tanto, también puede ser utilizado cuando alguna de las variables que conforman los individuos es de tipo nominal.

PAM. Partitioning around medoids (Mirkin, B., 2005)	KMEANS
1.- <i>Configuración inicial.</i> Elegido el número de clústeres, K , selecciona K entidades (<i>medoids</i>) $c_1, c_2, \dots, c_k \in I$ (muestra de individuos analizados) con un algoritmo especial llamado <i>Build</i> (básicamente, plantea un proceso de optimización en el que busca los K <i>medoids</i> tales que minimizan la disimilaridad entre ellos y los individuos pertenecientes a su grupo, con respecto a la disimilaridad a la que dichos individuos están de los pertenecientes a otros grupos). Asume que la lista inicial de clústeres S_k está vacía.	1.- Se seleccionan K centroides (inicialmente de forma aleatoria).
2.- <i>Actualización de clústeres.</i> Dados K <i>medoids</i> $c_k \in I$, se determinan los clústeres S'_k ($k = 1, \dots, K$) con la normal de la mínima distancia hasta los mismos aplicada a las disimiliadades $d(i,j)$, con $i, j \in I$.	2.- Cada individuo perteneciente a I es asignado a su centroide más cercano, tomando como indicador de disimilaridad el cuadrado de la distancia Euclidiana entre ambos.
3.- <i>Condición de parada.</i> Se valida si $S'=S$. Si es así, finaliza el clústering y $S=\{S_k\}$, $c=(c_k)$. En caso contrario, se cambia S por S' .	3.- Se actualizan los centroides de los nuevo clúster, igualándolos al promedio entre los individuos pertenecientes a cada uno de los mismos (dicho promedio hace referencia al de las variables que conforman un individuo, es decir, los centroides son vectores de medias de longitud p , donde p es el número de variables).
4.- <i>Actualización de los medoids.</i> Dados los S_k clusters, se determinan sus c_k medoids ($k=1, \dots, K$) y se vuelve al paso 2.	4.- Los individuos son reasignados a los nuevos centroides más cercanos.
	5.- Se repiten los pasos 3 y 4 hasta que las observaciones no son reasignadas o se alcanza el número máximo de iteraciones.

Tabla 3: Descripción de los algoritmos KMEANS y PAM.

2.1.3. Indicador de calidad del resultado obtenido: La silueta

El objetivo general perseguido por las técnicas de clústering consiste en identificar grupos compactos, es decir, clústeres con una similitud intra-clúster alta y una similitud inter-clúster baja. Esta idea intuitiva se formaliza, entre otros indicadores, en el concepto de *silueta* de un clúster (Kaufman, 1990). Su cálculo se realiza mediante la siguiente expresión:

$$S(i) = \frac{b_{ij} - a_i}{\max(a_i, b_{ij})}$$

Donde a_i es el resultado del promedio de las distancias entre individuos del clúster i y b_{ij} es el promedio de las distancias entre individuos del clúster i y del clúster j más cercano a i .

Las razones más significativas que justifican este modo de evaluar la calidad del clústering son:

- Es una medida aplicable a todos los métodos de clústering contemplados en el presente estudio.
- Está relacionada tanto con la similitud intra-clúster como la inter-clúster.
- Posee un valor normalizado fácilmente interpretable, que está comprendido entre -1, clústeres mal conformados, y 1, clústeres muy compactos.
- Su cálculo es sencillo. Además, la función *silhouette* (del paquete *cluster*) proporciona otro tipo de información de gran interés: el tamaño del clúster, el valor desilueta medio del clúster y el valor medio de silueta del conjunto de clústeres.

Es de importancia recalcar que no siempre se ha usado este indicador para evaluar las distintas soluciones, ya que el propósito de conformar agrupaciones compactas no siempre coincidía con el perseguido.

2.2. Proceso de calibración

Evolver es un algoritmo genético que utiliza un método de estado uniforme. Esto significa que sólo reemplaza un organismo en cada iteración, en lugar de una generación entera. Cuando crea un nuevo individuo, opta por dos padres de entre la población almacenada en un punto del proceso de búsqueda, cuyo tamaño es definido por el usuario. En tanto que utiliza un mecanismo de selección basado en la clasificación de individuos según el valor alcanzado por la función objetivo, existe mayor probabilidad de que sean escogidos como padres organismos con mejores valores asociados de la misma. Si bien, la curva de probabilidad de selección es más uniforme con respecto a calcularla proporcionalmente al valor de la función objetivo, tal y como lo hacen otros algoritmos, lo cual impide que individuos superiores dominen completamente la evolución desde el inicio. A continuación, el cruce realizado (se explica el correspondiente al método *receta básico*, que es el empleado en este trabajo) es de rutina uniforme. Esto significa que captura variables de uno de los padres aleatoriamente, de forma que no influye en el proceso de búsqueda la posición irrelevante en la que se encuentran las mismas. El método de solución de orden que lleva implementado introduce dichas variables capturadas anteriormente en el otro padre con el mismo orden en el que aparecen en el primero. Por otro lado, Evolver realiza mutaciones analizando cada variable de un individuo: se genera un número aleatorio entre 0 y 1 para cada una y si resulta ser menor o igual a la tasa de mutación definida por el usuario, entonces dicha variable es sustituida por un valor aleatorio.

Así pues, Evolver permite seleccionar el tamaño de la población almacenada, la tasa de cruce y mutación, así como la semilla origen de los procesos aleatorios comentados. En este trabajo se ha optado por probar combinaciones aleatorias de valores distintos de cada uno de dichos parámetros, que no distan en demasía en el caso de la tasa de cruce y mutación de los valores recomendados por el manual de Evolver.

Por otra parte, este software, tal y como ya se ha comentado, permite lanzar macros desde MS Excel, por lo que su conexión con Toolkit de EPANET es posible y, en consecuencia, también enfrentar el problema de la calibración de una red con el mismo. El proceso que implica esta fase de la modelación se puede visualizar en la **Figura 1**. Dada una hoja de cálculo abierta con Evolver correctamente configurada, en cada iteración, después de llevar a cabo las operaciones de cruce y mutación, se lanza una macro que adquiere los valores de las variables de decisión pertenecientes al nuevo individuo y los introduce en un archivo INP de EPANET, el cual contiene el modelo elaborado. Se lleva a cabo una simulación en EPANET y los resultados que arroja son registrados en la hoja de cálculo original, de forma que ya se está en disposición de evaluar la función objetivo. Dicho valor es leído por Evolver, a partir del cual determina la aptitud del individuo. Seguidamente, comprueba los criterios de parada: si se cumplen, detiene el proceso de búsqueda, en caso contrario, continua con la siguiente iteración.

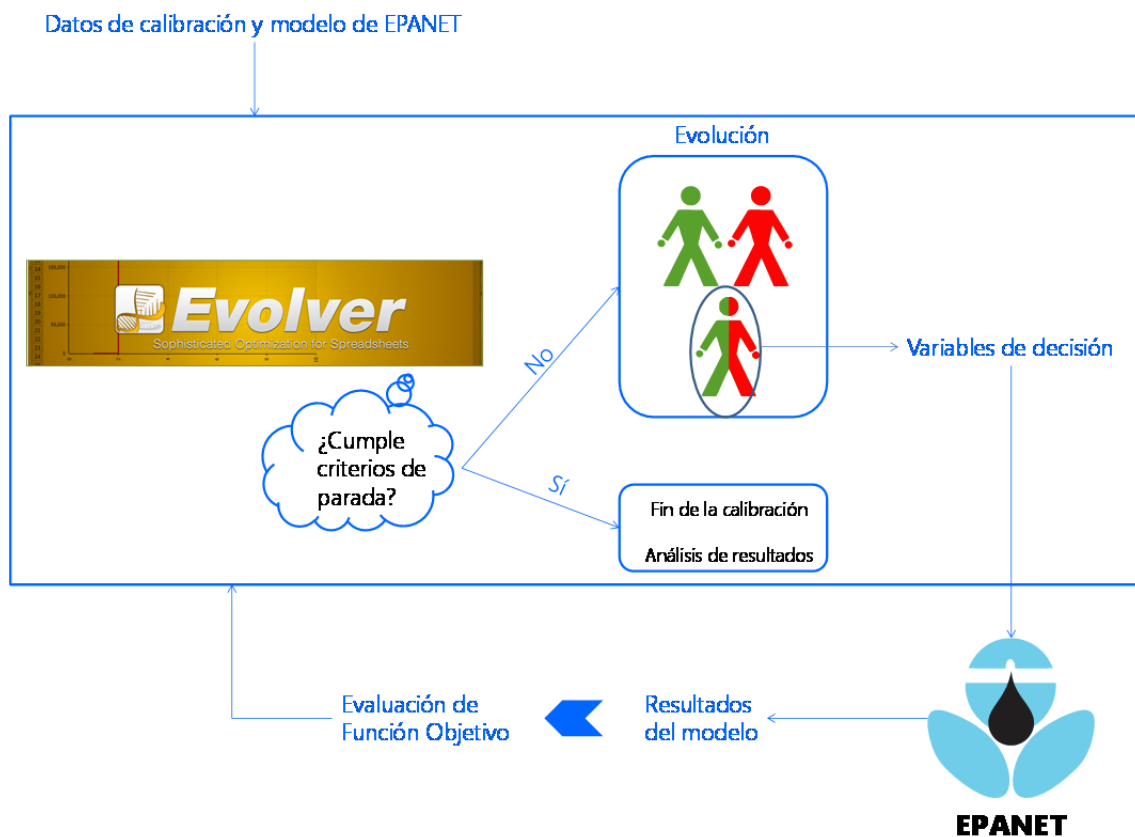


Figura 1: Esquema propuesto del proceso de calibración con Evolver para una red de abastecimiento de agua.

3. Caso de estudio de Nogoyá

3.1. Información general

3.1.1. Ubicación geográfica

El caso de estudio de la presente investigación es la red de abastecimiento de agua potable del municipio argentino de Nogoyá, cabecera del departamento con igual nombre y perteneciente a la provincia de Entre Ríos (**Figura 2**). Esta última integra junto a las provincias de Corrientes y Misiones la Mesopotamia argentina, formada por los ríos Paraná y Uruguay.



Figura 2: Ubicación geográfica de Nogoyá a escala estatal, provincial y departamental.

Está situado a unos 100 km al Sureste de la ciudad de Paraná, capital de la provincia de Entre Ríos, y próximo a la margen derecha del Arroyo Nogoyá.

3.1.2. Caracterización del medio físico natural

El término municipal de Nogoyá abarca una superficie de 11800 ha, mientras que su planta urbana cubre aproximadamente 500 ha. El terreno sobre el que se asienta es ondulado, con alturas que van desde 35 a 80 msnm, siendo la altura media de la planta urbana de 45 msnm.

El clima es de tipo templado húmedo de llanura, con una temperatura media anual de 18.4 °C, sin presentar situaciones extremas, y precipitaciones medias alrededor de 1100-1200 mm al año, con marcada variación interanual (ZAKER, 2008).

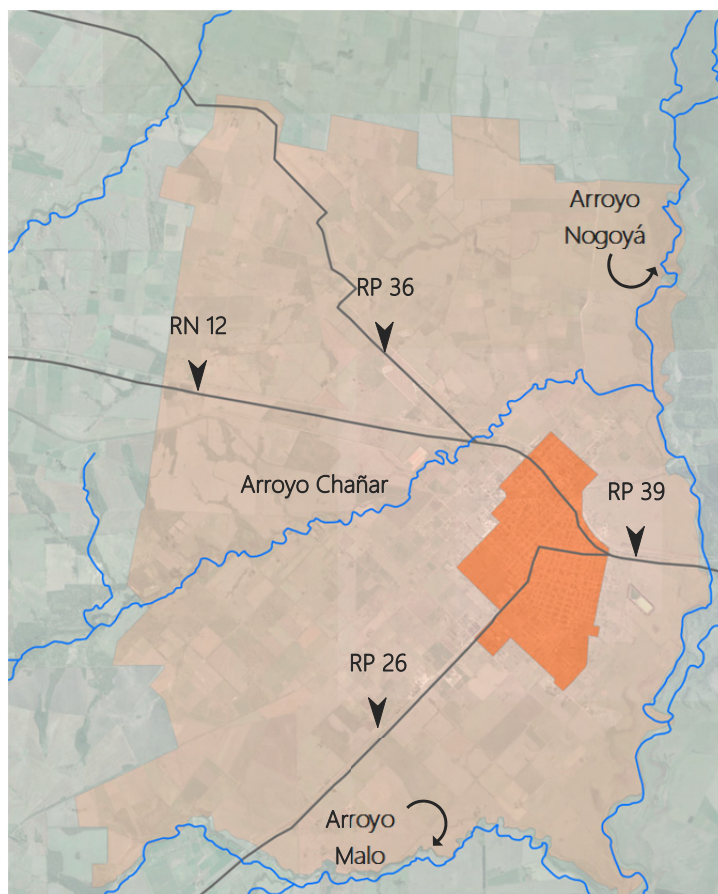


Figura 3: Ubicación de viales de comunicación y arroyos con respecto a la planta urbana de Nogoyá.

Los recursos hídricos superficiales los proporcionan diversos arroyos que recorren el ejido (**Figura 3**), limitado al Sur por el Arroyo Malo. La planta urbana queda enmarcada al Noreste por el Arroyo Chañar y al Oeste por el Arroyo Nogoyá, al cual desemboca el primero. Estos cursos de agua, receptores de los efluentes industriales y cloacales, son indicativos de las pendientes generales de escurrimiento (DSAPSBPER, 2012). Si bien, la fuente única de abastecimiento de agua potable es de origen subterráneo, procedente del llamado Sistema Acuífero Ituzaingó-Puelches⁴. Una revisión de pozos de los departamentos Paraná, Diamante y Nogoyá, llevada a cabo por Tomás *et al.* (1999), señala que los acuíferos confinados de la Formación Ituzaingó tienen un rendimiento que va entre 1.3 y 7.35 m³/h.m, siendo en general aguas bicarbonatadas sódicas de bajo contenido salino, el cual oscila entre 500 y 1200 ppm.

⁴ Con este nombre se designa al acuífero de aguas dulces que está incluido en la formación geológica de igual nombre y que ocupa gran parte del territorio de Corrientes, Entre Ríos y partes de la provincia de Buenos Aires

3.1.3. Caracterización socioeconómica

La infraestructura de comunicación principal es terrestre. En dirección Oeste-Este, está constituida por la Ruta Nacional (RN) N° 12, que conecta Nogoyá con las localidades de Crespo y Paraná, y la Ruta Provincial (RP) N° 39, que lo enlaza con Rosario de Tala. En dirección Norte-Sur, la RP N° 26 comunica el municipio con Victoria (**Figura 3**).

Nogoyá se encuentra en un área fuertemente ligada a la producción agropecuaria, siendo una región tradicionalmente tambera que dispone de la mayor capacidad instalada en industrias lácteas de la provincia de Entre Ríos (ZAHER, 2008). La fábrica más relevante en la localidad es "La Sibila", que procesa alrededor de un millón de litros diarios de leche. Está emplazada en el límite occidental de la planta urbana y dispone de sus propios pozos de aprovisionamiento de agua. A parte de ésta, existen otras 2 fábricas lácteas menores.

En la zona rural destacan el cultivo de cereales y oleaginosas. En la planta urbana de Nogoyá se ubican siete acopiadoras de cereales (DSAPSPER, 2012). Otro aporte a la agroindustria es la presencia de frigoríficos de ganado bovino (ZAHER, 2008).

Existen otros tipos de industria regional sita en el término municipal de Nogoyá, que elaboran pinturas, cerveza artesanal, pegamentos y productos metalúrgicos, así como talleres de distinta índole. Todas las actividades económicas que producen consumos de agua notables disponen de sus pozos privados para abastecerse.

3.1.4. Análisis demográfico

La población registrada en Nogoyá es de 23702 habitantes, según el último censo publicado por la Dirección General de Estadísticas y Censos de la Provincia de Entre Ríos en el año 2010. Ésta constituye un 60.7% de la población total del departamento de Nogoyá.

El anterior censo, realizado en 2001, inscribió en el municipio 21339 habitantes, por tanto, la tasa de crecimiento medio anual en la década comprendida entre ambos censos es de 1.06%, relativamente baja y propia de una localidad con dinámica poblacional consolidada. A su vez, en el último medio siglo se ha producido una emigración de la población rural dispersa en el ejido y región aledaña hacia la zona urbana, una tendencia que seguramente persistirá a futuro (DSAPSPER, 2012).

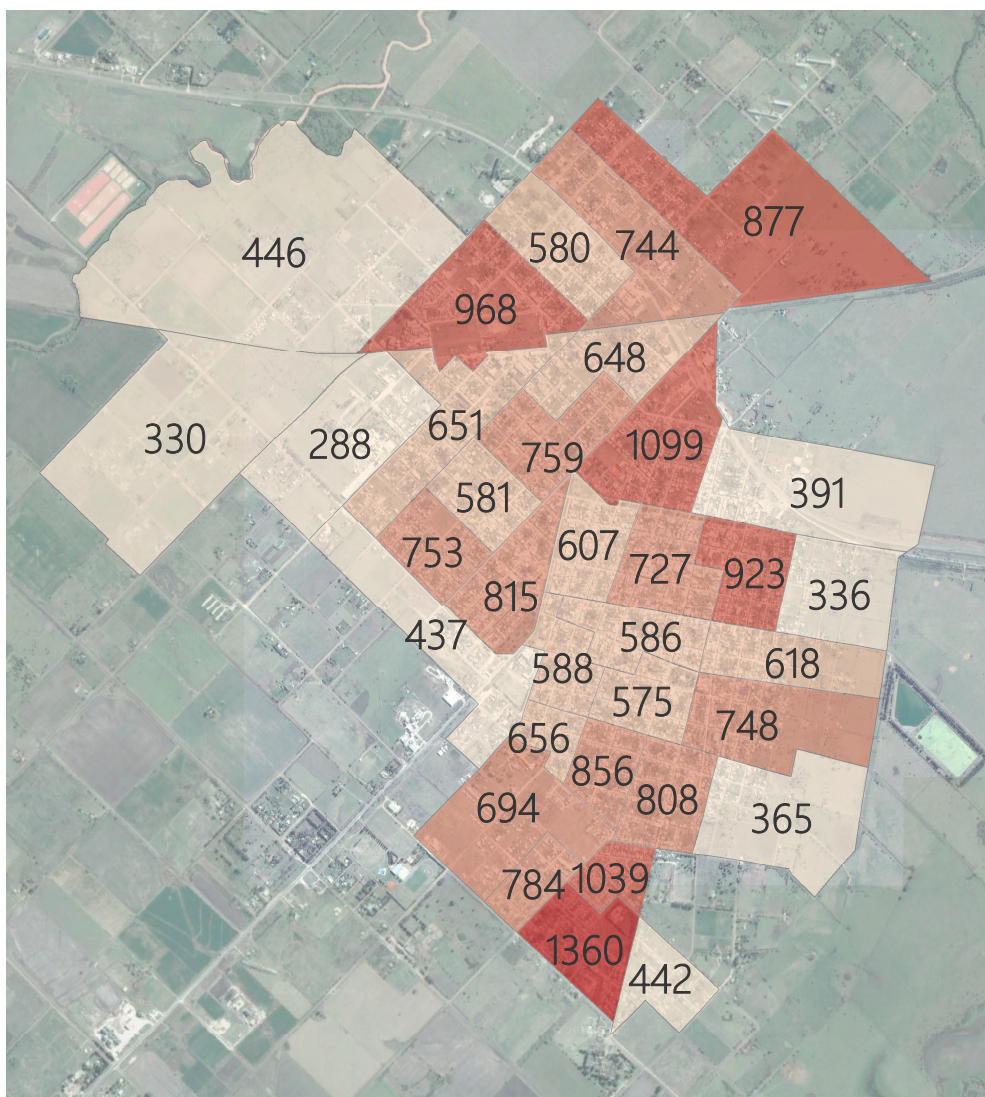


Figura 4: Número de habitantes totales en Nogoyá por radio censal según Censo 2010.

3.2. Descripción general del sistema de abastecimiento de agua

La localidad de Nogoyá dispone de una buena cobertura de servicios básicos, ya que el 87.9% de las viviendas particulares habitadas dispone de conexión satisfactoria a los mismos. El 98.9% de los hogares habitados alberga baño o letrina y el 98% se abastece de agua potable a través de la red pública (valores calculados a partir del Censo 2010, elaborado por la Dirección General de Estadísticas y Censos de la Provincia de Entre Ríos).

La Municipalidad (Ayuntamiento) de Nogoyá es el operador que gestiona las infraestructuras de distribución de agua potable y saneamiento. Las funciones técnicas de mantenimiento y operación son llevadas a cabo por el área de Obras Sanitarias, mientras que las funciones administrativas, de atención al usuario y cobro de tarifas quedan a cargo

de otras dependencias municipales, todas ellas adscritas a la Secretaría de Obras y Servicios Públicos.

El área de Obras Sanitarias está compuesta por el Jefe del Servicio y 30 operarios, distribuidos en cuadrillas. Éstas trabajan en turnos rotativos, de forma que la jornada ordinaria cubre 12 horas diarias a lo largo de todo el año. Por otro lado, el sistema de producción está bajo permanente vigilancia de un guardia.

Para la infraestructura de abastecimiento de agua potable en particular, no se evidencian déficits importantes de equipamiento para realizar operaciones regulares y básicas de mantenimiento. En contraste, no existe un conocimiento profundo del sistema desde el punto de vista técnico: el catastro no está actualizado ni en soporte digital fácilmente accesible ni funcional, existe información de importancia (por ejemplo, los partes de roturas) únicamente en soporte papel, se carece de micromedición o, en su defecto, de estudios sobre consumo fundamentados en mediciones que permitan hacer balances hídricos de cierta precisión, no se monitorean las presiones de servicio en red y no se dispone de dispositivos de medición del caudal inyectado.

La red de distribución está compuesta por 3 sectores (**Figura 5**):

- El **sector 1** es el de menor tamaño. Consta de una longitud total de tubería de 3.8 km y abastece a una población de 900 habitantes, aproximadamente. Está alimentado por el pozo 12 (características de la bomba en la **Tabla 5**), que impulsa agua al tanque 33 (características en la **Tabla 4**), conectado a la red.
- El **sector 2** tiene una longitud total de 11.4 km y proporciona agua a 3700 habitantes, aproximadamente. Se inyecta agua mediante bombeo directo procedente del pozo 10 (los datos de la electrobomba sumergida no están disponibles).
- El **sector 3** es el más grande. Abastece a 18600 habitantes y alcanza una longitud total de tubería de 81.9 km, aproximadamente. Dispone de dos puntos de alimentación:
 - El tanque principal del sistema, nro. 35 (características en la **Tabla 4**), al cual impulsan agua 7 pozos: el nro. 1, 2, 3, 4, 5, 6 y 8 (características de las bombas en la **Tabla 5**). Todos ellos acaban interconectándose en una única tubería que constituye la de subida al tanque.

El pozo 11. El mismo está conectado de forma tal que puede inyectar agua directamente a la red o impulsarla para el llenado del tanque 34, de características similares a las del tanque 33 (excepto por la cota de la solera, que está en torno a los 5 metros sobre el nivel del terreno). Respecto a la bomba instalada en el pozo, cabe decir que los datos facilitados no concuerdan con los proporcionados por la página web del proveedor. En el apartado 3.5 se explica de qué modo se ha dirimido esta disconformidad.

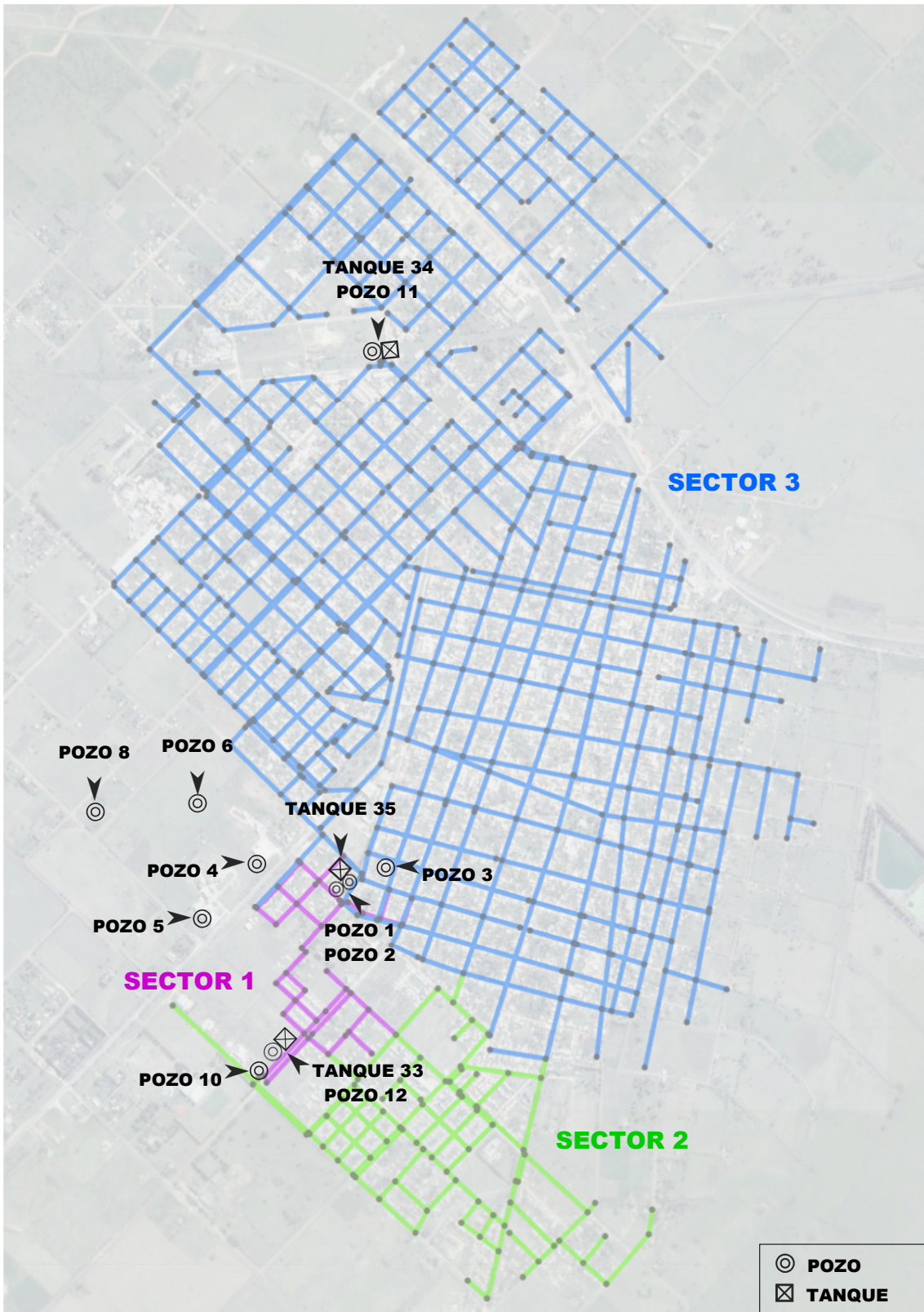


Figura 5: Esquema general de la red de abastecimiento del Municipio de Nogoyá.

Nº	Tipo	Material	Dimensiones	Cota solera	Nivel máx.	Nivel mín.
33	Cilíndrico	Poliestireno reforzado con fibra de vidrio	{ \varnothing 4,00 m; h = 7,00 m}	15,00 m s/NTN	22,00 m	15,00 m
35	Cilíndrico	Hormigón Armado	{ \varnothing 10,00 m; h = 5,00 m}	20,00 m s/NTN	25,40 m	20,00 m

Tabla 4: Características y dimensiones de los tanques 35 y 33.

Pozo Nº	Proveedor	Modelo-Serie	Pot. (HP)	Q (m ³ /h)	Hm (mca)
1	Rotor Pump	RP8S55-4K	25	50	65
2	Rotor Pump	RP8S64-4A	40	110	65
3	Rotor Pump	8RXSP 500-5	30	100	67
4	Rotor Pump	RP8S64-4A	40	110	69
5	Rotor Pump	RP8S64-4K	35	50	68
6	Rotor Pump	RP8S64-4K	35	80	70
8	Rotor Pump	RP8S64-4A	40	100	69
10	N/A	N/A	N/A	N/A	N/A
11	Rotor Pump	<i>E6</i>	<i>7.5</i>	50	55
12	Rotor Pump	RP8S50-6K	25	75	52

Tabla 5: Características técnicas de las electrobombas sumergibles instaladas en los distintos pozos (datos de octubre de 2014).

En la **Figura 6** puede observarse como la evolución temporal de la infraestructura ha dejado huellas fácilmente identificables, ya que en periodos localizados de desarrollo se ha empleado en su construcción un determinado material, así como el municipio se ha extendido por áreas concretas del ejido.

Las obras de la zona más antigua, ejecutadas por el ente gubernamental extinto Obras Sanitarias de la Nación, se llevaron a cabo en la década de 1940. Durante esta época se emplearon tuberías de fibrocemento. El esquema presentado de esta parte de la red se extrajo de la propia realidad, a partir de la ubicación de las válvulas de corte y de los partes de roturas, pues no existen planos. Esta zona tiene notables problemas de fugas y deficiente resistencia frente a presiones relativamente altas. Es por ello que dichas presiones se controlan con una válvula reductora (VRP) al pie del tanque principal, la cual disipa la energía correspondiente a más de un 50% de la cota total que proporciona el mismo. La función del resto de válvulas instaladas en el sistema (que suman alrededor de 200) es el de corte para operaciones de mantenimiento y reparación, por lo que están completamente abiertas en situación ordinaria.

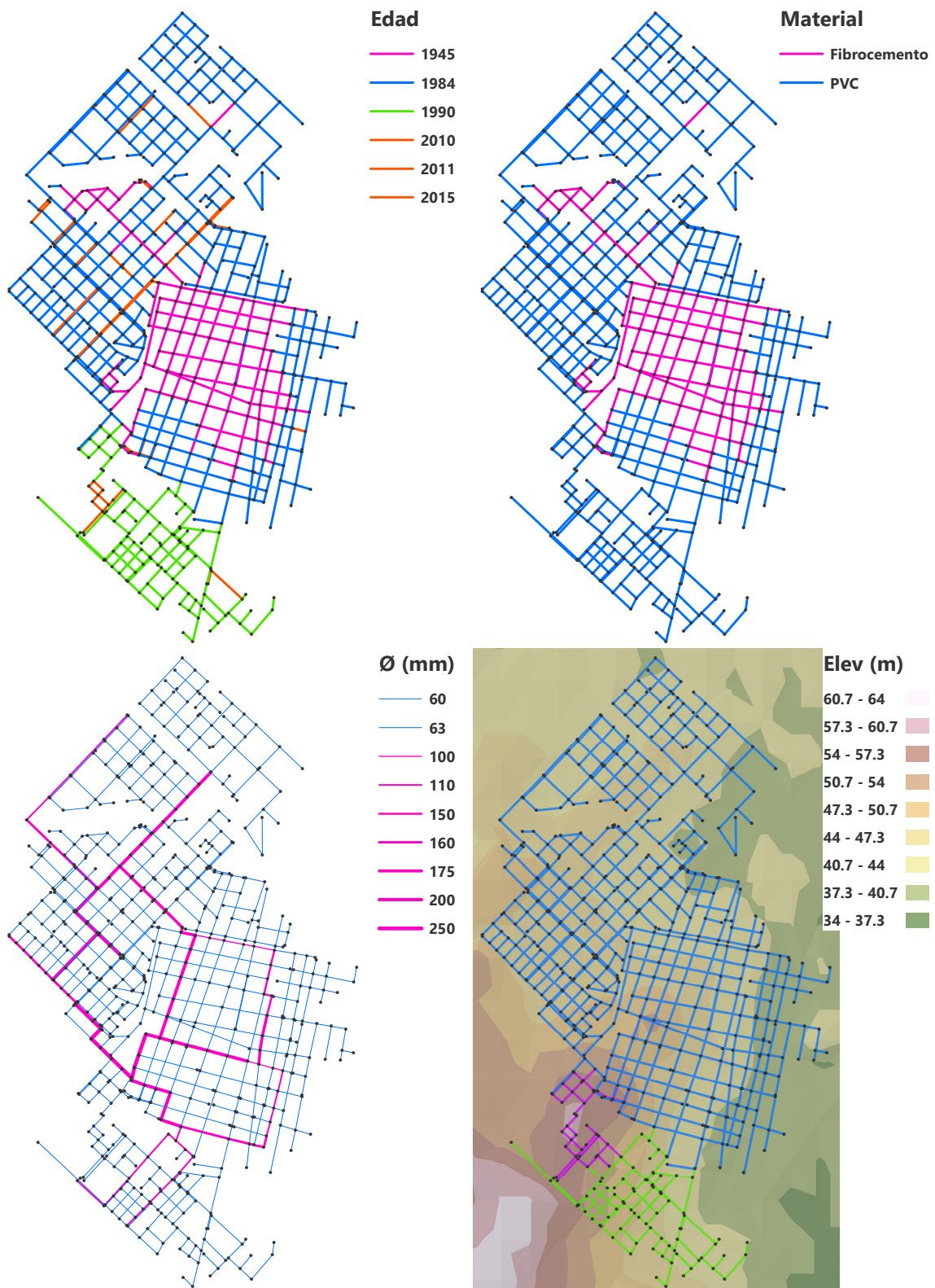


Figura 6: Edad, material y diámetro de la red y elevación del terreno en la zona de la planta urbana.

En la década de los 80 se produjo el mayor desarrollo: la red pasó de 2000 conexiones a 6000, aproximadamente. Ello hizo necesaria la construcción del pozo 11 y el tanque 34, que suponen un refuerzo de presión para las zonas dentro del sector 1 más alejadas del tanque principal. A partir de mediados de los años 90, la población empezó a extenderse hacia

áreas donde la cota del terreno incluso supera a la que se encuentra el tanque principal nro. 35. En virtud del gran recurso subterráneo disponible en la región, la solución adoptada fue crear dos nuevos puntos de producción, constituidos por el pozo 10 y el pozo 12 junto al tanque 33, que alimentan al sector 2 y 1, respectivamente. A lo largo de todo este periodo se emplearon tuberías de PVC.

La mayor parte de la red de distribución la componen tuberías de diámetro 60 mm (fibrocemento) y 63 mm (PVC). Por su lado, la red maestra incluye diámetros que van desde los 100 mm a los 250 mm y tuberías de los dos materiales presentes en el sistema. Su trazado describe grandes mallas, con lo que se garantiza la seguridad de poder ofrecer suministro frente a eventualidades en las que se vea afectada.

Para finalizar, añadir que en casi la totalidad de las parcela hay construida una única vivienda y ésta corresponde a una edificación de tipo unifamiliar de máximo 2 plantas. El 50% de las mismas, aproximadamente, disponen de tanque domiciliario. La instalación de fontanería en estos casos es normalmente de alimentación mixta: el grifo de la cocina procede de una conexión directa a red, mientras que el resto de la instalación está enlazada al tanque domiciliario.

3.3. Recopilación de información sobre la red de abastecimiento

3.3.1. Información que interviene en la elaboración del modelo

El listado que se presentó al equipo de trabajo con la información necesaria para elaborar el modelo está resumido en la **Tabla 6**. Este listado contiene la información de prioridad máxima. Otros ítems de menor importancia no se añadieron en pro de centrar los esfuerzos en aquello que era imprescindible.

Topología de la red	Plano de la red donde estén identificados los elementos de la misma
Conducciones	Diámetro
	Longitud
	Material
Bombas y/o estaciones de bombeo	<i>Equipo de bombeo</i>
	Potencia
	Rendimiento
	Horas de funcionamiento
	Niveles de aspiración
	Punto(s) de funcionamiento habitual/nominal (altura/caudal)

Elementos de regulación: válvulas	Tipo
	Diámetro
	Estado o consigna
Depósitos	Ubicación
	Tipo
	Dimensiones
	Cota solera
	Niveles máximo y mínimo
Puntos de consumo	Cota
	Tipo de consumo
	Consumo (medio+curva modulación, o medio+punta+valle)
Puntos de inyección	Cota+presión (en su caso)
	Caudal medio
	Modulación diaria y estacional
Abonados	Domicilio (ubicación georreferenciada o en plano)
	Disponibilidad de micromedición o no
	Consumo facturado (datos por sectores)
	Tipo de consumidor

Tabla 6: Listado de información con prioridad 1 acerca de la red de distribución de Nogoyá.

La información que finalmente fue posible conseguir queda resumida en la **Tabla 7**, en la que se especifica la fuente y el tipo de formato en el que se presentó. Tal y como puede observarse en la misma, el grueso de la información fue proporcionada por el jefe del servicio, el Ing. Daniel Rocca, ya que él constituía el principal enlace entre el CEGELAH y la Municipalidad de Nogoyá. El resto de información la facilitaron instituciones públicas, como la Municipalidad de María Grande, la Dirección de Estadísticas y Censos de la provincia y el Instituto Geográfico Nacional de la República Argentina, o fue obtenida a través de la web, como es el caso de los datos del Catastro de Nogoyá, extraídos del visor SIG habilitado por la empresa GEOMAT. La forma de comunicación con estas fuentes era principalmente a través de correo electrónico, excepto en el caso del trato con el Ing. Daniel Rocca, que era también presencial y vía teléfono móvil.

		Prioridad	Disponible	Características del formato en el que se ha obtenido la información y fuente principal
Topología de la red	Plano de la red donde estén identificados los elementos de la misma	1	Sí	El plano disponible está en AutoCAD. No dispone de georeferenciación ni de escala y las proporciones a lo largo del mismo no se mantienen. Fue proporcionado por el jefe de servicio del sistema, el Ing. Daniel Rocca.
	Conectividad entre elementos	1	Sí	En el plano de AutoCAD es posible identificar visualmente la conectividad entre elementos, pero el mismo no está

			editado haciendo referencia a objetos.		
	Por qué calles discurren las conducciones	1	Sí	Existen referencias en el plano de AutoCAD.	
	Localización de acometidas	1	No		
	Localización de válvulas de cierre y regulación	1	Sí	Dibujadas en el plano de AutoCAD.	
Conducciones	Diámetro	1	Sí	No existe un catastro de tuberías organizado en base de datos. Las características de las conducciones están implícitas en el plano de AutoCAD, excepto la longitud, que se ha obtenido de forma indirecta a partir de un plano de AutoCAD de la planta urbana georeferenciado y a escala, también proporcionado por el exjefe de servicio del sistema, el Ing. Daniel Rocca.	
	Longitud	1	Parcial		
	Material	1	Sí		
	<i>Pérdida de carga</i>				
		Pérdida de carga continua	2	No	
		Pérdida de carga localizada	2	No	
		Edad	2	Parcial	Obtención indirecta a partir de información secundaria extraída de conversaciones con el exjefe de servicio del sistema, el Ing. Daniel Rocca.
		Estado de conservación	2	No	
	<i>Equipo de bombeo</i>				
		Potencia	1	Parcial	Información del fabricante. No se conoce el tipo de bomba instalada en el pozo 10.
	Rendimiento	1	No		
	Horas de funcionamiento	1	Parcial	En el caso de los 7 pozos que bombean agua al tanque principal, registro del 2 al 12 de diciembre de 2014 en hoja de cálculo. En el resto de casos, a través de conversaciones con el exjefe de servicio del sistema, el Ing. Daniel Rocca.	
	Niveles de aspiración	1	No		
	Punto(s) de funcionamiento habitual/nominal (altura/caudal)	1	Parcial	Información del fabricante. No se conoce el tipo de bomba instalada en el pozo 10.	
	Curva característica	2	Parcial	Información del fabricante. No se conoce el tipo de bomba instalada en el pozo 10.	
	Conectividad	2	No		
Bombas y/o estaciones de bombeo		Modo de operación (arranque/paradas)	2	Parcial	Información aportada durante conversaciones mantenidas con el exjefe de servicio del sistema, el Ing. Daniel Rocca.
	<i>Conjunto de la estación</i>				
		Topología	2	No	
		Elementos de regulación	2	Parcial	Información aportada durante conversaciones mantenidas con el exjefe de servicio del sistema, el Ing. Daniel Rocca.
		Tanques de almacenamiento	2	No	
		Variación semanal y estacional en el modo de operación y regulación	2	Parcial	En el caso de los 7 pozos que bombean agua al tanque principal, registro del 2 al 12 de diciembre de 2014 en hoja de cálculo. En el resto de casos, a través de conversaciones con el exjefe de servicio del sistema, el Ing. Daniel Rocca.
		Dispositivos de medición de caudal	2	No	

	Altura de las bombas	2	No	
Elementos de regulación: válvulas	Tipo	1	Sí	Información aportada durante conversaciones mantenidas con el jefe de servicio del sistema, el Ing. Daniel Rocca.
	Diámetro	1	No	
	Estado o consigna	1	Sí	Información aportada durante conversaciones mantenidas con el jefe de servicio del sistema, el Ing. Daniel Rocca.
	Modelo	2	No	
	Características resistentes	2	No	
	Material	2	No	
	Edad	2	No	
	Estado de conservación	2	No	
Depósitos	Fabricante	3	No	
	Ubicación	1	Parcial	
	Tipo	1	Sí	
	Dimensiones	1	Sí	Información presentada en diferentes archivos de MS Excel y Word, proporcionada por el jefe de servicio del sistema, el Ing. Daniel Rocca.
	Cota solera	1	Sí	
	Niveles máximo y mínimo	1	Parcial	
	Evolución temporal de nivel	2	Parcial	
	Concentración de contaminantes	3	No	
Puntos de consumo	Cota	1	Parcial	Obtención indirecta a partir de las cotas identificadas en el plano de AutoCAD.
	Tipo de consumo	1	Parcial	Información aportada durante conversaciones mantenidas con el jefe de servicio del sistema, el Ing. Daniel Rocca.
	Consumo (medio+curva modulación o medio+punta+valle)	1	Parcial	Obtención indirecta a partir de información secundaria proporcionada por la Municipalidad de María Grande, la Dirección de Estadísticas y Censos de la Provincia de Entre Ríos y el visor SIG del catastro (elaborado por la empresa GEOMAT), a través de imágenes WMF, archivos DWG, hojas de cálculo, datos presentados en soporte papel, así como a través de descarga de archivos PDF que contenían información del catastro en el caso del visor SIG.
	Conectividad	2	No	
	Sector al que pertenece	2	Parcial	Obtención indirecta a partir de información secundaria procedente del visor SIG de GEOMAT y el plano de AutoCAD.
	Niveles de fugas en la zona de influencia del punto	2	No	
	Abonados que se abastecen desde el punto	2	Parcial	Obtención indirecta a partir de información secundaria proporcionada por el visor SIG de GEOMAT y la Dirección de Estadísticas y Censos de la Provincia de Entre Ríos, sumada al plano de AutoCAD.
	Tipo de suministro	2	Sí	Información aportada durante conversaciones mantenidas con el jefe de servicio del sistema, el Ing. Daniel Rocca.
Puntos de inyección	Cota+presión (en su caso)	1	Parcial	Obtención indirecta a partir de las cotas identificadas en el plano de AutoCAD y de
	Caudal medio	1	Parcial	Mediciones de caudal inyectado del día 2 al 12 de diciembre de 2014 en el punto de
	Modulación diaria y	1	Parcial	

	estacional			producción más importante (generado por los 7 pozos que bombean al tanque principal), registrado en hoja de cálculo.
	Históricos de caudal	2	No	
	Concentración de reactivos	2	No	
	Domicilio (ubicación georreferenciada o en plano)	1	No	
	Disponibilidad de micromedición o no	1	Sí	Información aportada durante conversaciones mantenidas con el exjefe de servicio del sistema, el Ing. Daniel Rocca.
	Consumo facturado (datos por sectores)	1	No	
Abonados	Tipo de consumidor	1	Parcial	Información aportada durante conversaciones mantenidas con el exjefe de servicio del sistema, el Ing. Daniel Rocca.
	Modulación diaria y estacional	2	No	
	Calibre, marca, edad y estado del contador divisionario	3	No	

Tabla 7: Listado completo de información requerida a la contraparte para la elaboración del modelo. Descripción de prioridad, disponibilidad, formato y fuente principal de información para cada ítem.

Es destacable que información relevante no llegó a estar disponible hasta mediados de mayo de 2015, así como hubo información que se pensó podría obtenerse en un principio y finalmente no fue posible. Ejemplo del primer caso, los registros de micromedición del municipio de María Grande, y del segundo, la distribución de acometidas o los datos que arrojaban los 27 micromedidores instalados de forma experimental en algunas viviendas de Nogoyá. Por otro lado, el formato en el que se podía acceder a cierta información requería grandes cargas de trabajo para que pudiera ser aprovechada, tal y como ocurrió con los datos del Catastro, que tuvieron que ser descargados de un visor web no preparado para ello. A lo anterior, se suma que la información procedía de distintas fuentes, en consecuencia, para poder cruzarla ha sido necesario llevar a cabo la ardua tarea de uniformar identificadores.

Cabe añadir que el hecho de que la estancia se realizará en la ciudad de Paraná, donde estaba ubicado el CEGELAH, supuso una dificultad para realizar visitas a campo con asiduidad y, finalmente, sólo fue posible desplazarse 2 veces hasta Nogoyá en 3 meses de estancia. Por otro lado, los integrantes del equipo de trabajo únicamente pudimos reunirnos alrededor de 3 veces por mes. Todo ello ocasionó que la transferencia de información, la resolución de dudas, así como la gestión y aprobación de propuestas para mejorar la cantidad o calidad de los datos, fueran discontinuas y espaciadas en el tiempo.

3.3.2. Información que interviene en la fase de calibración

3.3.2.1. Caudales

Únicamente fue posible medir el caudal en la tubería de subida al tanque principal nro. 35. Para ello se empleó el siguiente equipo:

- i. Caudalímetro intrusivo ODIN, modelo TIL-3300 HT, con válvula auxiliar para la instalación e interfaz óptica para la toma de lectura manual. Sin data logger.

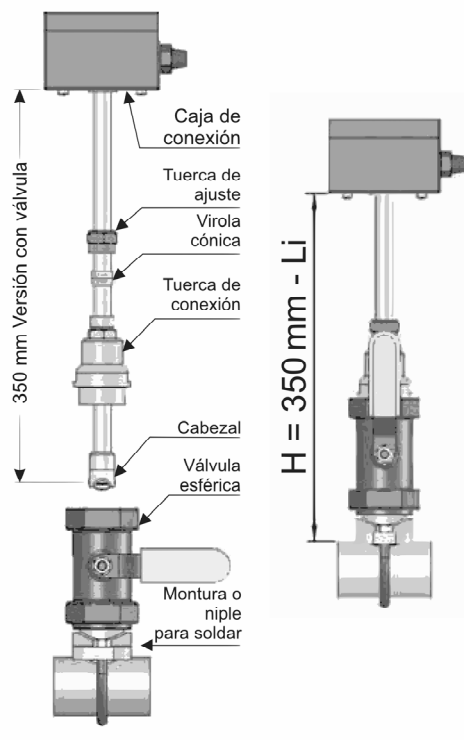


Figura 7: Esquema del caudalímetro intrusivo y acotación de la distancia H que marca la profundidad de inserción para que el eje del rotor quede a una profundidad de $1/3$ del radio. Fuente: Adaptado del Catálogo de ODIN.

Previo vaciado del tanque, se procedió a su instalación⁵ el día 1 de diciembre de 2014. Teniendo en cuenta que el diámetro interior de la tubería de subida al tanque principal nro. 35 es igual a 303.2 mm, la profundidad de inserción de la turbina se ajustó tomando una distancia H de 28.9 cm ($L_i = 61$ cm) entre la pared exterior de la tubería y la cara inferior de

⁵ La tubería ya se encontraba perforada, puesto que hubo un intento anterior de puesta en marcha. Éste falló debido a que una de las hélices de la turbina estaba fracturada y, a causa de ello, el sensor no proporcionaba lecturas coherentes, por lo que tuvo que ser reparado.

la caja de conexión, tal y como se muestra en la **Figura 7**, de forma que el eje del rotor quedó a una profundidad de $1/3$ del radio.



Figura 8: Instantánea tomada durante la instalación del caudalímetro intrusivo. Fuente: propia.

No se contaba con data logger, por lo que el registro se realizó de forma manual. Al inicio de cada hora se apuntaba el volumen acumulado medido, el nivel del tanque y la presión aguas abajo de la válvula reductora localizada en cabecera de red. El registro abarca el periodo comprendido entre el día 2 a las 00:00 horas y el día 13 a las 00:00 horas de diciembre de 2014.

Cabe añadir que se intentó llevar a cabo mediciones en otros puntos de inyección a red a través de una cooperación con el Ente Regulador de los Recursos Termales de la Provincia de Entre Ríos, que posee caudalímetros no intrusivos, si bien, no fue posible.

3.3.2.2. Presiones

Las mediciones disponibles proceden del estudio realizado para el informe *"Diagnóstico sobre el uso de energía eléctrica en los abastecimientos de agua y saneamiento. Experiencia piloto en el Municipio de Nogoyá."*, fruto de un convenio entre la Secretaria de Energía de la Provincia de Entre Ríos y la UADER, publicado el año 2013. Se llevaron a cabo con los equipos descritos a continuación:

- i. 2 transmisores de presión Siemens D-76181 Karlsruhe, modelo STIRNAS P serie ZD, cod. 7MF1580-1EA10, serial AZB/B3114873 y AZB/B3114875.
- ii. 2 Data Logger Novus, modelo LOGBOX-AA, serie 11011266 y 11011230.

iii. 2 Interface Óptica Novus, modelo IR-LINK 3USB, serie 1107548 y 1107395.

Las sesiones realizadas se extendieron desde julio de 2012 a marzo de 2013. Se organizaron en dos vueltas: la primera estaba comprendida en la época invierno-otoño y la segunda en la época primavera-verano.



ID	Domicilio
P1	17 de Agosto 1967
P2	Avda. Marchini y Bv. Alberdi
P3	Bv. España 1346
P4	Bv. España y 25 de Mayo
P5	F. Gerald y Q. y Taboada 876
P6	J.J. Paso y Marchini
P7	Avda. Marchini y Colon
P8	Avda. Marchini 1582
P9	Q. y Taboada 876
P10	Ruta 12 y Victoria
P11	Yrigoyene/Parana y S. Oriz

Figura 9: Ubicación en la planta urbana de los puntos donde se han llevado a cabo mediciones de presión. En la tabla de la izquierda aparece la dirección del domicilio facilitada a la que corresponde cada uno de dichos puntos.

Los criterios de instalación del sensor de presión y recolección de datos fueron:

- Acometidas domiciliarias con conexión directas a red de distribución.
- Sensor programado para toma de datos cada 10 minutos.

- Duración total del registro de 72 horas en sesiones de la época invierno-otoño y primeras de la época primavera-verano. De 48 horas en posteriores.
- Datos complementarios: dirección del domicilio, fecha y hora de inicio y fin de la toma de datos, información climática. También se especifica el diámetro de la tubería de distribución a la que conecta la acometida donde se instala el medidor.
- Encuestas para caracterización del consumo en cada domicilio.
- Archivo plano con el registro completo.

En la **Figura 9** se puede consultar la ubicación en la planta urbana de los puntos donde se han observado presiones. En el apartado 3.5 se analizarán con detenimiento los datos recabados en cada uno de los mismos.

3.4. Post-proceso de información

3.4.1. Tratamiento de información que interviene en la fase de modelación

3.4.1.1. Sistema de Información Geográfica (SIG) para el sistema de abastecimiento de agua de Nogoyá

Si consultamos la definición que proporciona el Instituto Geográfico Nacional español sobre el término Sistema de Información Geográfica (SIG), cabe destacar para este estudio dos acepciones:

- “El SIG como proyecto, cada una de las realizaciones prácticas, de las implementaciones existentes”. La definición completa es: “Conjunto integrado de medios y métodos informáticos, capaz de recoger, verificar, almacenar, gestionar, actualizar, manipular, recuperar, transformar, analizar, mostrar y transferir datos espacialmente referidos a la Tierra” (Bosque J.,1992).
- “El SIG como software, es decir, los programas y aplicaciones de un proyecto SIG”.

Así pues, en este trabajo por necesidad se ha iniciado un proyecto SIG, con el objetivo de crear un repositorio de gran potencial con la información disponible del sistema de abastecimiento de agua de Nogoyá, en tanto que dicha información queda georeferenciada y está presentada en un formato útil que puede ser empleado como input de herramientas imprescindibles para la modelación. El software utilizado en el mismo ha sido ArcGIS.

El presente apartado se estructura en dos secciones: en la primera se define el sistema de coordenadas empleado como marco de referencia y en la segunda se exponen los distintos bloques de trabajo, concernientes a distintas temáticas (trama urbana, topología, elevación del terreno y censo), en los que se podría dividir los avances realizados en el proyecto SIG iniciado.

3.4.1.1.1. Sistema de Coordenadas

Un sistema de coordenadas es el marco de referencia que permite ubicar información espacialmente sobre la Tierra. Queda definido por:

- El datum, que es el conjunto de los parámetros a y b del elipsoide (modelo del geode, el cual es definido por la Geodesia, ciencia que estudia la forma y tamaño de la Tierra, como la superficie en la que todos los puntos experimentan la misma atracción gravitatoria, siendo ésta equivalente a la observada al nivel del mar), las coordenadas geográficas $[\lambda, \omega]$ del punto fundamental (en el que el elipsoide y el geode son tangentes) y la dirección que define el Norte (imagen izquierda de la **Figura 10**). Hoy en día, con la fuerte presencia de los Sistemas de Posicionamiento Global (GPS), se busca que los datum tengan validez para todo el planeta. Ello se consigue con un parámetro más: la distancia del centro del elipsoide con respecto al centro de masas de la Tierra (SIGMUR, 2005).

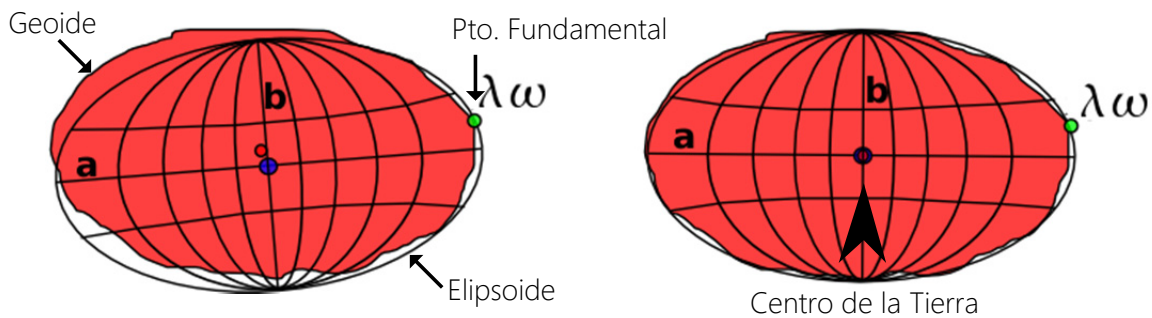


Figura 10: Conjunto de parámetros que definen un datum. Fuente: Adaptado de SIGMUR, 2005.

- El marco de medición y unidad de medida asociada, que puede ser:
 - Geográfico, el cual equivale a un sistema de coordenadas esféricas (propio de espacios euclídeos tridimensionales) con centro coincidente con el de la Tierra, que usa las coordenadas angulares, latitud (Norte y Sur) y longitud (Este y Oeste), medidas tradicionalmente en grados decimales.
 - Planimétrico, que requiere de la definición de una proyección cartográfica (la cual permite pasar de un marco de medición geográfico a uno planimétrico) y usa las coordenadas cartesianas $[x, y]$, también llamadas planas, expresadas habitualmente en metros.

El sistema de coordenadas utilizado en el presente trabajo está condicionado por la información disponible. El AutoCAD proporcionado con la trama urbana a escala y georeferenciado se encuentra en el Sistema de Coordenadas Planimétrico definido por el Marco de Referencia Geodésico Nacional POSGAR 94 (Posiciones Geodésicas Argentinas), que emplea el datum WGS-84 (el habitualmente usado por los GPS), y la proyección de Gauss-Krüger, el cual ha sido seleccionado como marco de referencia para el presente proyecto SIG. Dicho sistema no está entre los incluidos en la lista de predeterminados de

ArcGIS, por lo que, dado que Nogoyá se encuentra en la Faja 5, se ha creado introduciendo los datos que se especifican en la **Tabla 8**.

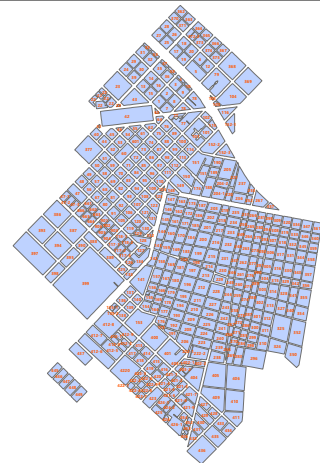
Proyección	
Falsa dirección Este:	5500000.0
Falsa dirección Norte:	0.0
Meridiano central:	-60.0
Factor de escala:	1.0
Origen de latitudes:	-90.0
<hr/>	
Unidad lineal:	Metro (1.0)
<hr/>	
Sistema de Coordenadas Geográfico GCS_WGS_1984	
Unidad angular:	Grados decimales (0.0174532925199433)
Primer meridiano:	Greenwich (0.0)
Datum:	D_WGS_1984
Esferoide:	WGS_1984
Semieje mayor:	6378137.0
Semieje menor:	6356752.314245179

Tabla 8: Definición del Sistema de Coordenadas Planimétrico POSGAR 94 y proyección Gauss-Krüger para la Faja 5.

3.4.1.1.2. BLOQUE 1: Trama urbana

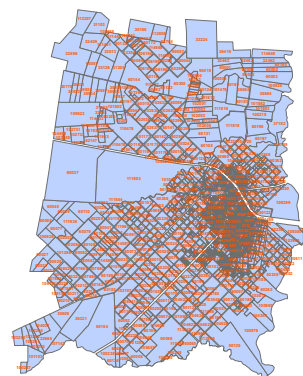
El propósito perseguido en este primer bloque es transvasar toda la información en AutoCAD sobre la planta urbana de Nogoyá al SIG. El resultado lo constituyen las capas descritas en la **Tabla 9**. La información de partida se enumera en la **Tabla 10**.

Nombre de la capa	Descripción
105_104_708_SpatialJoinEditTableText_Man	Capa de polígonos que incluye las manzanas y su correspondiente identificador según la información proporcionada en AutoCAD.



205_204_710_SpatialJoin_Parc

Capa de polígonos que incluye las parcelas y su correspondiente identificador según la información proporcionada en AutoCAD.



302_301_Edit_TramaUrbana

Capa de polígonos que delimita la planta urbana consolidada.

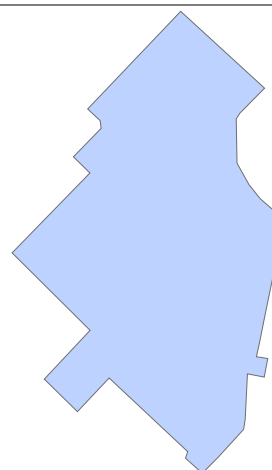


Tabla 9: Resultados obtenidos en el Bloque 1 del SIG, Trama Urbana.

ID	Nombre	Descripción
DTU03	DTU03_PlanoAEscalaNogoya.dwg	Plano en AutoCAD de la planta urbana y el término municipal de Nogoyá, a escala y georeferenciado (en el Sistema de Coordenadas Planimétrico definido por POSGAR 94 y la proyección Gauss-Krüger).

Tabla 10: Información de partida para el desarrollo del Bloque 1 del SIG, Trama Urbana.

Los pasos para conseguir las capas resultado son análogos, por lo que se procede a explicar únicamente los seguidos en la obtención de la capa *105_104_708_SpatialJoinEditTableText_Manz*, que quedan recogidos en el diagrama de flujo de la **Figura 11**.

Paso 1: Guardar las capas de AutoCAD requeridas como capa SIG

El archivo AutoCAD *DTU03* contiene en sus capas *Polymanz* y *R_Polymanz* las manzanas dibujadas como polilíneas. Se visualizan en ArcGIS únicamente dichas entidades, para seleccionarlas y posteriormente guardarlas como capa SIG de polilíneas, con nombre *101_DTU03_ManzPolilin*.

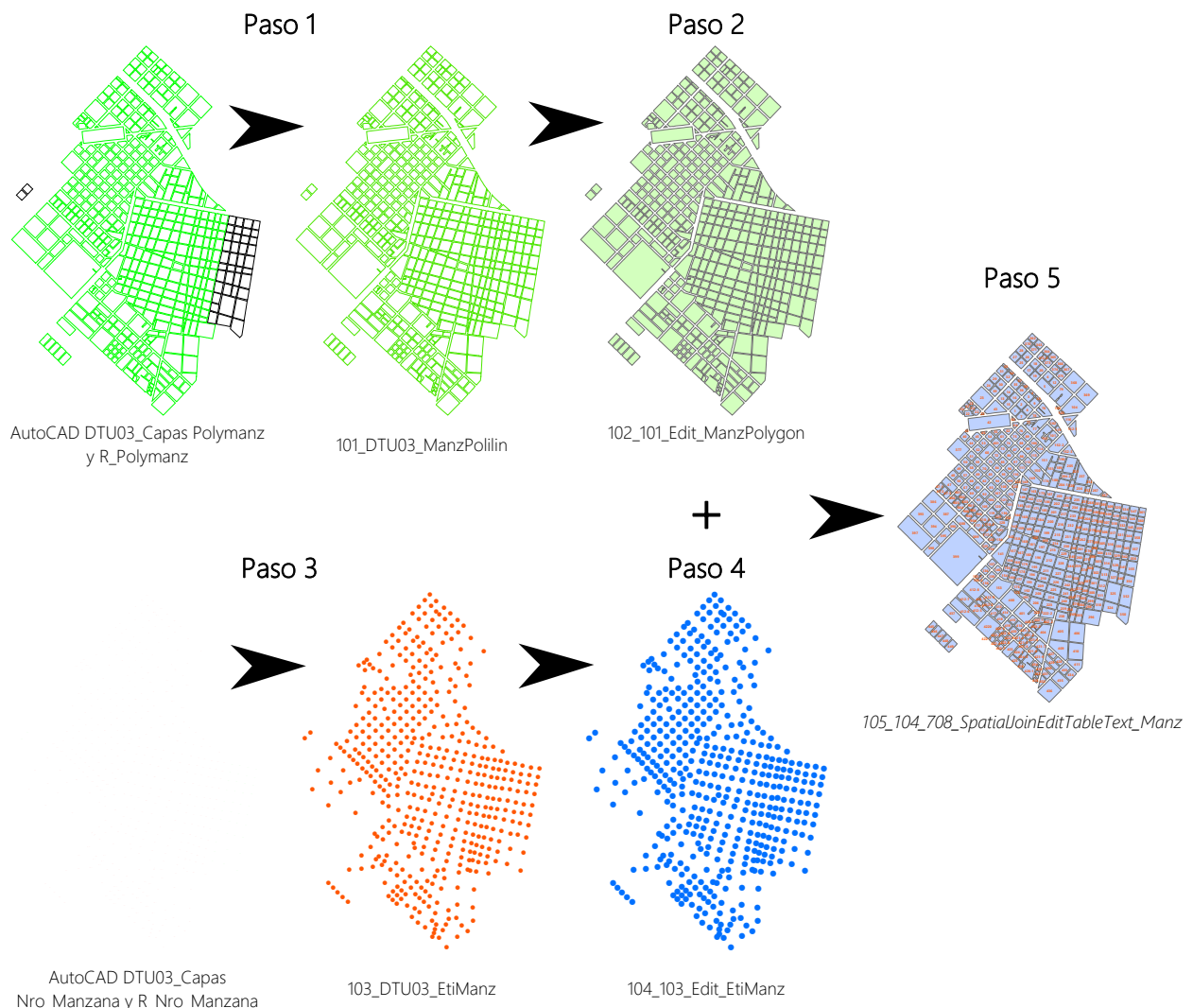


Figura 11: Diagrama de flujo para la obtención de la capa 105_104_708_SpatialJoinEditTableText_Maniz.

Paso 2: Edición de las manzanas como polígonos a partir de la capa 101_DTU03_ManizPolilin

En primer lugar, se crea una capa de polígonos, con nombre *102_101_Edit_ManizPolygon*, asociándole el sistema de coordenadas del proyecto. A continuación, se inicia el modo *Edición*, se seleccionan todas las entidades de la capa *101_DTU03_ManizPolilin* y se unen en una sola con la herramienta de edición *Merge*. Seguidamente, se crean las nuevas entidades polígono en la capa *102_101_Edit_ManizPolygon* con la herramienta *Auto Complete Polygon*, según la secuencia: clic en algún punto de la entidad única polilínea de *101_DTU03_ManizPolilin* → botón derecho → seleccionar *Replace Sketch*.

Paso 3: Transformación de las etiquetas con el identificador de manzana en AutoCAD a punto

Se carga en la tabla de contenidos por separado las entidades *Anotaciones* del archivo AutoCAD *DTU03*. Mediante selección de las capas *Nro_Manzana* y *R_Nro_Manzana*, se visualizan únicamente las etiquetas que contienen los identificadores de manzana. El uso de la herramienta *Conversion Tools > To Shapefile > Feature Class To Shapefile* permite la

transformación de las etiquetas a puntos. Su aplicación proporciona como resultado la capa *103_DTU03_EtiManz*.

Paso 4: Edición de la capa 103_DTU03_EtiManz

Normalmente algunos de los puntos resultantes del anterior paso no quedarán contenidos dentro de su polígono correspondiente, hecho que no permite trasladar espacialmente a los mismos la información que contienen dichos puntos (Figura 12). Es por ello que habrá que corregir mediante edición la posición de los puntos conflictivos. Estos últimos pueden ser identificados seleccionando espacialmente los polígonos de *102_101_Edit_Manzipolygon* según la capa de puntos que contienen la información de las etiquetas.

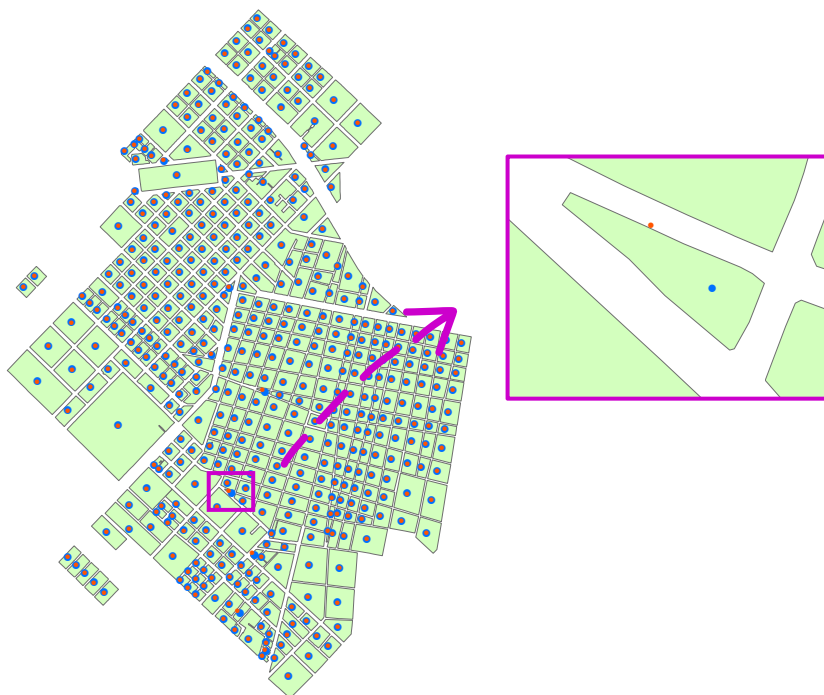


Figura 12: Ejemplo de punto conflictivo.

Se guardan las capas que van a ser editadas para asegurar el trabajo realizado hasta el momento. Es por ello que la capa resultado es *104_103_Edit_EtiManz*.

Paso 5: Unir espacialmente las capas 102_101_Edit_Manzipolygon y 104_103_Edit_EtiManz

Esta operación se puede llevar a cabo, por ejemplo, mediante la herramienta *Analysis Tools > Overlay > Spatial Join*. Como resultado se obtiene la capa *105_104_708_SpatialJoinEditTableText_Manzipolygon*.

3.4.1.1.3. BLOQUE 2: Topología de la red

Este bloque está especialmente subordinado a la información disponible. El AutoCAD donde aparece dibujada la red no está a escala, siquiera guarda las proporciones. Además, la edición del mismo no se realizó en su día haciendo referencia a objetos, por lo que los vértices de las líneas y los nudos no coinciden especialmente entre sí. Pese a que se realizaron intentos para aprovechar el esquema de la red proporcionado, con herramientas como *Adjustment spatial* del modo Edición de ArcGis, finalmente se decidió que era más rápido crear la red desde cero. El resultado conseguido se materializa en las capas incluidas en la **Tabla 11**. Los archivos de los que se alimenta este bloque se pueden consultar en la **Tabla 12**.

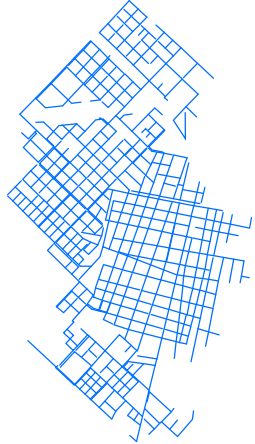

Nombre de la capa	Descripción
512_510_GHydr_TubORI	<p>Capa de líneas que incluye todas las tuberías de la red y la información de importancia asociada: identificador, longitud aproximada, tipo de material y su código, fecha de construcción aproximada y su código, sector al que pertenecen e interconexión con los nudos de la capa 513_511_GHydr_NudORI.</p> 
513_511_GHydr_NudORI	<p>Capa de puntos que contiene los nudos de la red. Todos ellos corresponden a interconexiones entre 2 o más tuberías con distintas direcciones, excepto los identificados como 345902 y 315902, que se crearon desafortunadamente en tuberías que seguían un trazado curvo.</p> 

Tabla 11: Resultados obtenidos en el Bloque 2 del SIG, Topología de la Red.

ID	Nombre	Descripción
DR01	DR01_EsquemaRedAbastSinEscala_laydel.dwg	Plano de la planta urbana sin escala y sin georeferenciar, que incluye el esquema de la red y su topología, así como los elementos principales de la misma identificados.
DTU03	DTU03_PlanoAEscalaNogoya.dwg	Plano de la planta urbana y el término municipal de

Nogoyá, a escala y georeferenciado (en el Sistema de Coordenadas Planimétrico definido por POSGAR 94 y la proyección Gauss-Krüger), que contiene una capa donde están dibujados los ejes de las calles.

Tabla 12: Información de partida para el desarrollo del Bloque 2 del SIG, Topología de la Red.

Es importante advertir que el trazado de las tuberías sigue el eje de las calles en general, ya que se usó como base una de las capas del AutoCAD DTU03. En consecuencia, el resultado obtenido con respecto a la longitud y trazado de las tuberías es aproximado, si bien se considera válido para el objetivo que debe cumplir: ser dato de entrada en la elaboración del modelo estratégico de la red. Su empleo para cualquier otro propósito está desaconsejado.

Los pasos para conseguir las capas resultado se enumeran a continuación y quedan resumidos en el diagrama de flujo de la **Figura 13**.



Figura 13: Diagrama de flujo para la obtención de las capas 512_510_GHydr_TubORI y 513_511_GHydr_NudORI.

Paso 1: Guardar la capa de AutoCAD usada de base en la edición de las tuberías como capa SIG

El archivo AutoCAD *DTU03* contiene en su capa *Calles_Ejes* los ejes de las calles dibujados como polilíneas. Se visualizan en ArcGIS únicamente dichas entidades, para seleccionarlas y posteriormente guardarlas como capa SIG de polilíneas, con nombre *501_DTU03_EjeCalles*.

Paso 2: Seleccionar por localización espacial los ejes comprendidos en la planta urbana y crear puntos en los vértices de las líneas resultantes

En primer lugar, se seleccionan por localización los ejes comprendidos dentro de la planta urbana delimitada por *302_301_Edit_TramaUrbana* y se guardan en una nueva capa con nombre *502_501_Select_OrigenTuberias*. A continuación, a través de la herramienta *Data Management Tools > Features > Feature Vertices To Points*, se generan puntos en los vértices de las líneas que conforman los ejes seleccionados y se registran en la capa *503_502_FeaturesVerticesToPoints_OrigenNudos*.

Paso 3: Edición de las capas 502_501_Select_OrigenTuberias y 503_502_FeaturesVerticesToPoints_OrigenNudos

Se abre el modo edición y se modifican simultáneamente las capas *502_501_Select_OrigenTuberias* y *503_502_FeaturesVerticesToPoints_OrigenNudos* para que reproduzcan de la forma más fiel posible el esquema proporcionado en el AutoCAD *DR01*. En cuanto a criterios de edición, se destaca:

1. Los finales de línea de los distintos ramales se hacen llegar hasta el término de la calle por la que discurren, es decir, algunos de ellos acaban antes y tienen menor longitud según el esquema dibujado en el AutoCAD *DR01*, si bien no se intenta emular en este sentido ya que no se dispone del dato exacto de longitud.
2. Normalmente, las cañerías maestras se crean sobre los propios ejes de las calles y se retranqueaban las tuberías de distribución de menor diámetro.

Apuntar, que sólo el ramal que empieza en la *Calle Carlos Gardel* (ubicada cerca de la cabecera del sector 2 de la red) no ha sido añadido, debido a que dicha calle y las manzanas circundantes, empleadas en otros casos como guía, no estaban dibujadas en el AutoCAD *DTU03*.

Por último, una vez finalizada la edición, se usa la herramienta *Data Management Tools > General > Delete Identical* para borrar los nudos redundantes allí donde coinciden los vértices de 2 o más ejes.

Paso 4: Crear las relaciones topológicas con GISRed y comprobación del resultado con la asistencia de GISRed y ArcGIS

La extensión de ArcGIS 3.X GISRed 1.0 se aplica para crear las relaciones topológicas entre los elementos nudo y tubería, así como para comprobar que éstas son correctas. El dato de entrada es la capa de tuberías editada resultado del anterior paso. Las operaciones realizadas son muy básicas:

1. Crear un nuevo proyecto en GISRed.
2. Importar (opción incluida en *Scenario*, apartado perteneciente al menú principal) la capa SIG de tuberías editada con ArcGis (para que se active dicha opción debe estar seleccionada la capa de tuberías que se crea por defecto con el proyecto).
3. Mediante la herramienta *Edit Model > Data Verification*, GISRed 1.0 genera las relaciones topológicas y una capa de puntos coherente con las mismas.
4. En la carpeta definida en el directorio de trabajo del proyecto están guardadas las capas resultado de las operaciones llevadas a cabo en GISRed. Se abren con ArcGIS y se compara visualmente la capa de puntos creada a través de GISRed y la editada originalmente con ArcGIS. Se puede usar la selección por localización para ver las diferencias entre una y otra y que queden rápidamente destacados errores cometidos, aunque se aconseja realizar una inspección detallada.

Otra estrategia para identificar errores con ArcGIS es:

1. Guardar las entidades tubería resultado del paso 3 como una nueva capa (*0_compr_512*).
2. Abrir el modo edición para la misma y mediante la herramienta *Merge* crear una única polilínea que contiene la totalidad de las tuberías. Se guardan los cambios y se cierra el modo edición.
3. Aplicar la herramienta *Data Management Tools > Features > Feature Vertices To Points* a la polilínea resultado del anterior paso, seleccionando la opción *Dangle* en el último despegable, con la que se obtiene una capa de puntos coincidentes con los nudos final de línea presentes en la red (*0_compr_512_n*). De forma visual se puede comprobar si existe algún final de línea no presente en la realidad.

Todos los procesos que implican este paso, dan como resultado las capas *513_511_GHydr_NudORI* y *512_510_GHydr_TubORI*.

Paso 5: Añadir información de interés. Nomenclatura para los identificadores de nudo y tubería.

En las capas finales se añadieron los campos:

512_510_GHydr_TubORI:

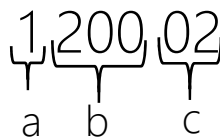
1. *Sector*: Sector de la red al que pertenece la tubería.
2. *DC_ID*: Identificador único de la tubería.
3. *NODE1*: Nudo en el que empieza la tubería.
4. *NODE2*: Nudo en el que finaliza la tubería.
5. *LENGTH*: Longitud de la tubería en metros.
6. *DIAMETER*: Diámetro de la tubería en milímetros.
7. *Fecha_Cons*: Fecha aproximada de construcción de la tubería.
8. *Desc_Mat*: Material del que está hecho la tubería.

513_511_GHydr_NudORI:

1. *Sector*: Sector al que pertenece el nudo.
2. *DC_ID*: Identificador único del nudo.

La información se introdujo con la ayuda de la *Calculadora de Campo*, previa selección de los elementos cuyas características se querían completar.

La nomenclatura implantada (con la asistencia de MS Excel) en los identificadores únicos de cada elemento sigue la siguiente estructura:



- a) Número del sector al que pertenece el elemento.
- b) Número de orden del tipo de elemento dentro del sector.
- c) Tipo de elemento. Los códigos a este respecto empleados en el proyecto son:
 - i. 01: Nudos de consumo.
 - ii. 02: Nudos de conexión.
 - iii. 03: Tuberías simples.
 - iv. 06: Electrobomba.

En la capa *512_510_GHydr_TubORI* y *513_511_GHydr_NudORI* sólo existen tuberías simples y nudos de conexión, respectivamente.

3.4.1.1.4. BLOQUE 3: Elevación del terreno

La información de la que parte el presente bloque es particularmente dispersa. Tal y como puede consultarse en la **Tabla 14**, proviene de tres fuentes distintas. Los resultados alcanzados se exponen en la **Tabla 13**.

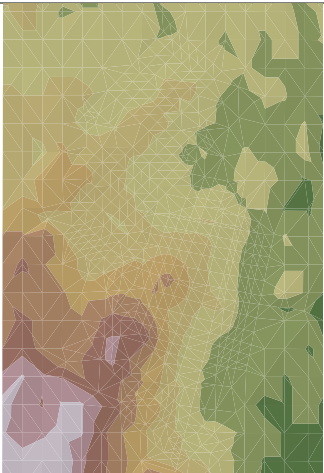
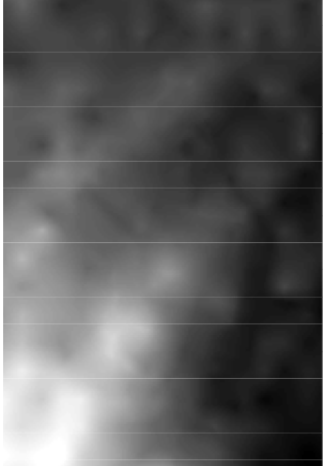
Nombre de la capa	Descripción	
617_616_TIN	Red Irregular de Triángulos (TIN) que modela el terreno sobre el que se asienta la red de abastecimiento de agua de Nogoyá.	
618_617_mde	Modelo Digital del Terreno (MDT) sobre el que se asienta la red de abastecimiento de agua de Nogoyá.	

Tabla 13: Resultados obtenidos en el Bloque 3 del SIG, Elevación del Terreno.

ID	Nombre	Descripción
DT01	DT01_Cotas.dwg	Plano con las cotas a la altura de las tapas de los pozos de registro de la red de saneamiento. Estás se ubican en el centro de los cruces de calle. Proporcionada por el exjefe de servicio, el Ing. Daniel Rocca.
DT02	DT02_CotasComplementariasVisorGeomat.xlsx	Cotas de los cruces de calle no cubiertos por la información contenida en DT01, extraídas del visor SIG web habilitado por la empresa GEOMAT.
DT03	DT03_336010.img	MDE, hoja 336010, publicado por el Instituto Geográfico Nacional de la República Argentina (http://ign.gob.ar/node/987).
DT05	DT05_Grilla_Select336010.dwg	Área de trabajo y cuadrícula de selección de cotas complementarias fuera de la planta urbana de DT03.

Tabla 14: Información de partida para el desarrollo del Bloque 3 del SIG, Elevación del Terreno.

Los pasos para conseguir las capas resultado se enumeran a continuación y quedan resumidos en el diagrama de flujo de la **Figura 13**.

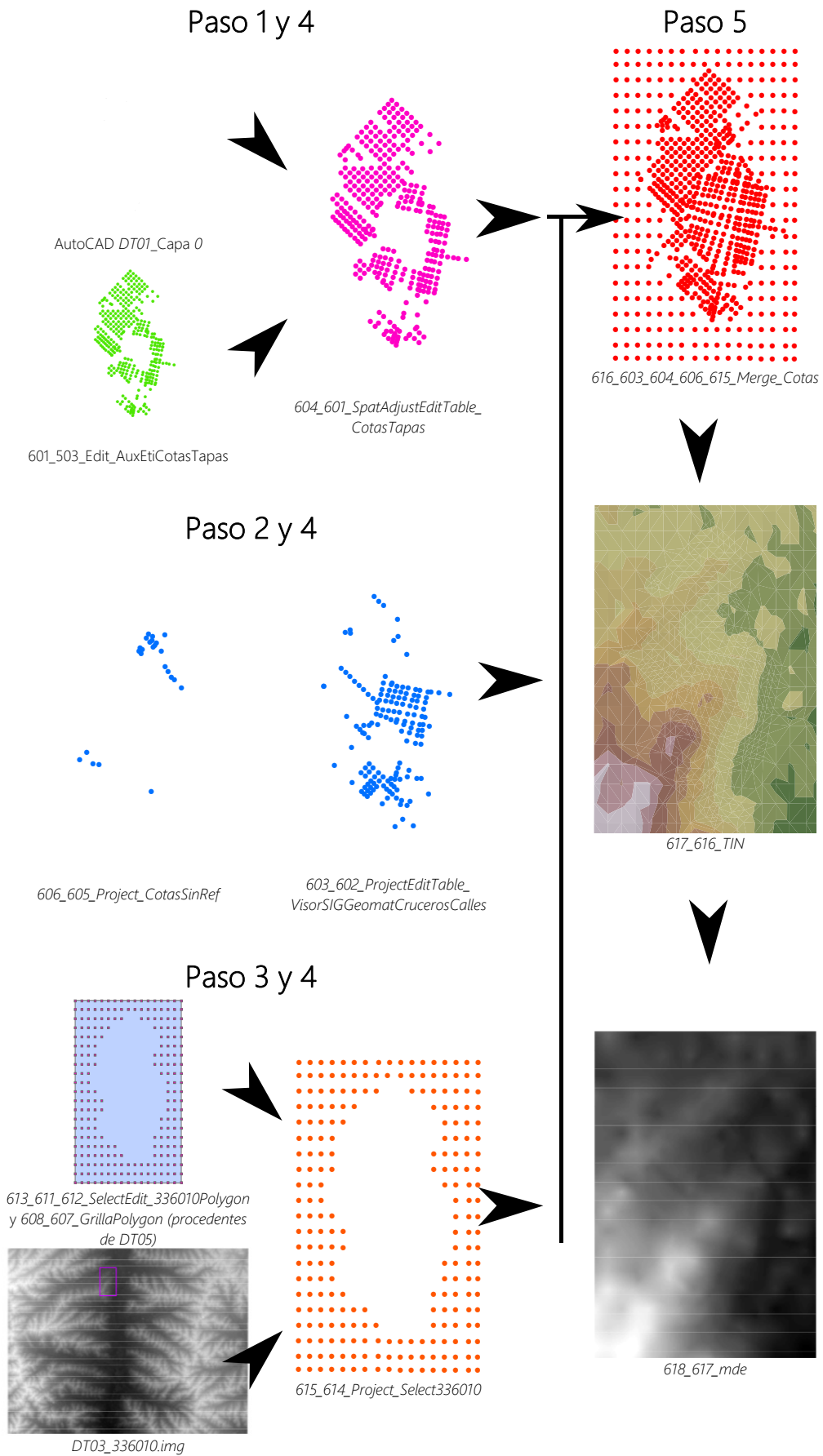


Figura 14: Diagrama de flujo para la obtención de 617_616_TIN y 618_617_mde.

Paso 1: Tratamiento de la información procedente de DT01

En primer lugar, se procede a la edición del archivo *modificaciones de Cloacal 2012-2014.dwg*, facilitado por el jefe del servicio, el Ing. Daniel Rocca, del que parte *DT01*. Dicha edición se realiza a través del propio AutoCAD. Su propósito es mantener en *DT01* únicamente las etiquetas con las cotas de las tapas de los pozos de registro de la red de saneamiento, dibujadas en la capa 0 del AutoCAD.

El archivo *modificaciones de Cloacal 2012-2014.dwg* no está a escala ni georeferenciado, al igual que pasaba con el esquema de la red de abastecimiento. A causa de ello, se decidió crear una guía de los cruces de calle en ArcGis, sólo allí donde se ubican espacialmente las cotas de las tapas de los pozos de registro que aparecen en *modificaciones de Cloacal 2012-2014.dwg*. Como base, se utiliza el archivo *503_502_FeaturesVerticesToPoints_OrigenNudos*, cuya modificación se guarda como *601_503_Edit_AuxEtiCotasTapas*, el cual sólo contiene los cruces de calle con cota en *DT01*, y como *602_503_Edit_AuxTodasCotasCruceCalles*, que contiene todos los que están contenidos dentro de la planta urbana consolidada.

A continuación, las etiquetas de *DT01* se transforman a capa SIG de puntos mediante la herramienta *Conversion Tools > To Shapefile > Feature Class To Shapefile*. Se abre el modo edición para la misma y, con la asistencia de *Spatial Adjustment*, se hace coincidir la ubicación de los puntos que contienen la información de la cota con la de los pertenecientes a *601_503_Edit_AuxEtiCotasTapas*. El resultado de esta operación se materializa en *604_601_SpatAdjustEditTable_CotasTapas*.

Paso 2: Tratamiento de la información procedente de DT02

Tal y como puede observarse en la **Figura 14**, las cotas disponibles en *modificaciones de Cloacal 2012-2014.dwg* no cubren toda el área de estudio. La empresa GEOMAT, cuyo principal encargo por parte de la Municipalidad es actualizar el Catastro de Nogoyá, publica en su visor SIG web el muestreo de una cantidad considerable de puntos donde se ha medido la cota. Es destacable que todos ellos se localizan dentro de la planta urbana consolidada. No quedó otra alternativa que extraer las cotas faltantes, que se creyeron complementarias a las presentes en *modificaciones de Cloacal 2012-2014.dwg*, de esta fuente. La operación fue ineficiente, ya que el visor no dispone de ninguna opción para realizar descargas de forma masiva, y de resultados imprecisos, pues las herramientas habilitadas para la consulta sólo permiten conseguir las coordenadas geográficas de los puntos de cota conocida de forma aproximada. Esta información se fue almacenando en una hoja de cálculo, con nombre *DT02_CotasComplementariasVisorGeomat.xlsx*. Cabe añadir que los puntos se separaron en dos grupos: aquellos que están sobre los cruces de calle (identificados como tal en

la descripción de la información ofrecida por GEOMAT) y los que no tienen una referencia espacial clara.

Seguidamente, se cargan los datos de *DT02* en ArcGis, a través de *File > Add Data > Add XY Data* en dos capas distintas, según los grupos creados. Sólo las coordenadas de los puntos de cota en cruce de calle se hacen coincidir con la de los correspondientes en *602_503_Edit_AuxTodasCotasCruceCalles*. Se asume sin remedio la inexactitud de las coordenadas del grupo restante. Las capas resultado son *603_602_ProjectEditTable_VisorSIGGeomatCruceCalles* y *606_605_Project_CotasSinRef*, respectivamente, ambas proyectadas en el sistema de coordenadas definido.

Paso 3: Tratamiento de la información procedente de DT03

En tanto que algunas partes de la red seguían quedando fuera del área en la que se disponía de puntos de cota conocida, se decidió terminar de cubrir la zona de trabajo con información procedente del Modelo Digital de Elevaciones (MDE-Ar) publicado por el Instituto Geográfico Nacional de la República Argentina. La hoja en la que se ubica Nogoyá es la número 336010 y se le ha adjudicado el nombre de *DT03_336010.img*. Es conveniente remarcar que la cota de un punto muestreado en *DT01* o *DT02* es menor que la correspondiente en el MDE-Ar, por lo que usar la información de este último supone quedarse del lado de la seguridad.

Las operaciones realizadas son:

1. Dibujar en AutoCAD la delimitación de la zona de trabajo. Sobre la misma elaborar una plantilla de selección conformada por recuadros de 60x60 m para extraer cotas del MDE-Ar cada 200 metros. Se toma esta distancia para que la densidad de puntos procedentes del MDE-Ar no condicionen excesivamente la interpolación entre cotas. El archivo resultado se nombra *DT05_Grilla_Select336010.dwg*.
2. Transformar las polilíneas dibujadas en *DT05* en capas de polígonos, tal y como se hizo con manzanas y parcelas en el BLOQUE 1. En el caso de la plantilla para la selección de cotas contenidas en *DT03* se suprime el área cubierta por *DT01* y *DT02*, editada en *611_ZonaCotaMuestreada*. Se obtienen así *608_607_GrillaPolygon* y *613_611_612_SelectEdit_336010Polygon*.
3. Recortar el ráster *DT03* según *608_607_GrillaPolygon*, con lo que se obtiene un nuevo ráster, el cual se denomina *609_DT03_608*.
4. Convertir *609_DT03_608* en una capa de puntos, designada como *610_609_RasterToPoint*, mediante *Conversion Tools > From Raster > Raster To Point*.
5. Seleccionar por localización los puntos de *610_609_RasterToPoint* contenidos en *613_611_612_SelectEdit_336010Polygon*. En aquellos casos en los que un

recuadro capta una pareja de puntos, se ha borrado uno de ellos en el modo edición. Se consigue así la capa *614_610_613_Select336010*.

6. Se proyecta *614_610_613_Select336010* al sistema de coordenadas del proyecto, con lo que se obtiene *615_614_Project_Select336010*.

Paso 4: Crear campos para la identificación de la procedencia de la información

Antes de unir los resultados finales alcanzados en los pasos anteriores (cuyos nombres están resaltados en negrita), se añaden los siguientes campos en cada uno de ellos:

- *Cota*: se crea un nuevo campo para la cota, de características uniformes en todas las capas.
- *Cod1* y *Desc1*: hace referencia a si la cota se ha obtenido mediante medición (*Zona muestreada*) o no.
- *Cod2* y *Desc2*: identifica la fuente de información.

Paso 5: Unión de las capa resultado de anteriores pasos e interpolación de cotas

Se procede a la unión de las capas:

- *604_601_SpatAdjustEditTable_CotasTapas*
- *603_602_ProjectEditTable_VisorSIGGeomatCrucerosCalles*
- *606_605_Project_CotasSinRef*
- *615_614_Project_Select336010*

Esto se lleva a cabo mediante la herramienta Data Management Tools > General > Merge, que devuelve la capa de nombre *616_603_604_606_615_Merge_Cotas*. Dicha capa es usada como dato de entrada para la conformación del TIN llamado *617_616_TIN*, resultado de la aplicación de la herramienta *3D Analyst Tools > Data Management > TIN > Create TIN* (opciones por defecto y sistema de coordenadas del proyecto). El MDE *618_617_mde* se ha obtenido a partir de *617_616_TIN* gracias a la herramienta *3D Analyst Tools > Conversion > From TIN > TIN To Raster* (opciones por defecto).

Paso 6: Dotar de cota a los nudos de la red

Para asociar la cota que corresponde a los nudos de la red, contenidos en la capa *513_511_GHydr_NudORI*, se emplea la herramienta *3D Analyst Tools > 3D Features > Add Z Information* (opciones por defecto), introduciendo como dato de entrada *617_616_TIN*.

El TIN elaborado pretende ser un modelo de la cota a pie de calle. Según conversaciones con el jefe de servicio, el Ing. Daniel Rocca, las tuberías se encuentran enterradas aproximadamente 1-1.2 metros en el caso de la zona antigua y 0.8-1 metros en

posteriores ampliaciones. En consecuencia, se ha decidido restar 1 metro a la cota procedente de 617_616_TIN como primera aproximación.

3.4.1.1.5. BLOQUE 4: Censo 2010 (Población y vivienda)

Los datos del Censo 2010, elaborado por la Dirección de Estadísticas y Censos de la Provincia de Entre Ríos, se considera una buena fuente de información a partir de la cual corroborar órdenes de magnitud en cuanto a densidad de viviendas y cantidad de consumo por zonas, en este último caso realizando estimaciones según la distribución de la población. Es por ello que se desarrolló el presente bloque, cuyo resultado final aparece en la **Tabla 15**. En la **Tabla 16** puede consultarse la información de partida.

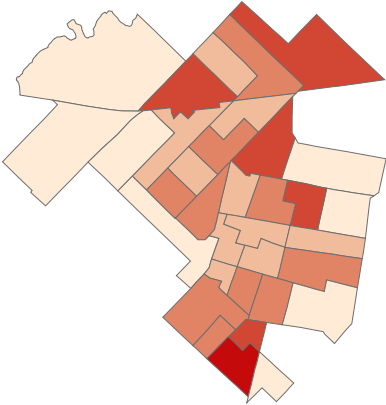
Nombre de la capa	Descripción
802_RadioCensalCenso2010	<p>Capa de polígonos en la que están delimitados los radios censales. Incluye información sobre vivienda y población totales por radio censal⁶, así como ratio de habitantes por vivienda promedio.</p> 

Tabla 15: Resultado obtenido en el Bloque 4 del SIG, Censo 2010.

ID	Nombre	Descripción
DC01	DC01_DelimitacionRCyFNogoya.pdf	Plano de la planta urbana de Nogoyá donde están demarcados radios censales y fracciones ⁷ .
DC02	DC02_1P-FR.xls	Total de población según sexo por departamento, por área de gobierno local, por fracción y radio censal.
DC03	DC03_6H-FR.xls	Hogares habitados según procedencia del agua para beber y cocinar por departamento, área de gobierno local, fracción y radio censal.

Tabla 16: Información de partida del Bloque 4, Censo 2010.

⁶ Unidad espacial mínima en los estudios realizados por la Dirección de Estadísticas y Censos de la Provincia de Entre Ríos.

⁷ Una fracción incluye varios radios censales.

El Bloque 4 conlleva una carga de trabajo menor respecto a bloques anteriores. Únicamente se dibujaron los polígonos que representan fracciones (*801_FraccionesCenso2010*) y radios censales (*802_RadioCensalCenso2010*), con la ayuda de herramientas del modo Edición como *Auto Complete Polygon* o *Trace* (que permite crear segmento siguiendo el trazado de elementos ya existentes), y seguidamente se introdujo la información de interés que correspondía a cada radio censal de forma manual. Para esto último fue necesario añadir los siguientes campos:

- *Radio_cens*: Identificador del radio censal.
- *Fraccion*: Identificador de la fracción.
- *Pob_tot*: Población total por radio censal.
- *Viv_tot*: Viviendas totales (independientemente de la procedencia del agua para beber y cocinar) por radio censal.
- *habxvivRC*: Ratio habitantes-vivienda por radio censal, calculado a partir de los campos *Pob_tot* y *Viv_tot*.

3.4.1.2. Tratamiento de la información que concierne a la estimación del consumo de Nogoyá

3.4.1.2.1. Post-proceso y análisis de datos de María Grande

El mayor hándicap a la hora de enfrentarse a la modelación del sistema de abastecimiento de Nogoyá proviene de la inexistencia de micromedición, es decir, no existen datos basados en mediciones sobre el consumo de agua. Todos los usuarios están acogidos a la cuota fija según zonas predefinidas por la Municipalidad (la delimitación de éstas pueden consultarse en la Ordenanza N° 891), pese a que está contemplada la tarificación por servicio medido.

La solución propuesta a este inconveniente aprovecha los datos de micromedición procedentes de un municipio de la misma zona que Nogoyá, María Grande. La premisa de partida es que resulta esperable que propiedades inmobiliarias (parcelas) de características similares, en distintas poblaciones de la misma región, posean como grupo una variabilidad parecida en relación al consumo. Se ha de destacar que, tanto en Nogoyá como en María Grande, dado que son municipios de tamaño pequeño-mediano, prácticamente la totalidad de las parcelas comprenden una vivienda y que la edificación predominante no supera las 2 alturas.

La idea principal expuesta en el anterior párrafo conduce en primer lugar a encontrar mediante procesos de clústering dichos grupos de parcelas similares según las siguientes características: superficie total de la parcela, superficie construida y habitantes por vivienda promedio. El resultado obtenido sirve como datos de entrada para el entrenamiento de un Soport Vector Maching (SVM), que permite clasificar las parcelas de Nogoyá según los clústeres definidos en María Grande, tal y como se muestra en el sub-apartado 3.4.1.2.1.3. En

los apartados 3.5 y 3.6 se desarrollan los pasos restantes que conforman la estrategia propuesta en relación a la modelación del consumo, definida en el apartado 2.

3.4.1.2.1.1. Fuentes de Información

María Grande es una población de 7694 habitantes, según el Censo 2010, que se encuentra a unos 80 km al Noroeste de Nogoyá y que previsiblemente tiene sus mismas costumbres de consumo de agua. La información que facilitó la Municipalidad de dicha población puede consultarse en la **Tabla 17**, la cual permite determinar la superficie total y construida de cada parcela perteneciente a la misma, así como su consumo asociado.

Identificador	Nombre	Descripción	Formato
DMG01	DMG01_ConsumoAnualMedioAsoCaracParcMariaGrande.xlsx	Registro de consumo medio anual por partida en el año 2013. La misma está descrita por su identificador, superficie total y superficie construida.	Papel
DMG02	DMG02_PlantaUrbanaParcelas.dwg	Archivo de AutoCAD de la planta urbana de María Grande, donde están dibujadas calles, manzanas y parcelas, junto a su identificador.	DWG
DMG03	DMG03_RegistroConsumosBimensuales2013.xls	Registro de las lecturas bimensuales de contadores realizadas en 2013 por partida.	XLS
DMG04	DMG04_RegistroConsumosBimensuales2014.xls	Registro de las lecturas bimensuales de contadores realizadas en 2014 por partida.	XLS

Tabla 17: Información proporcionada por la Municipalidad de María Grande.

Por otro lado, la información empleada en el cálculo del ratio habitantes por vivienda y radio censal proviene del Censo 2010. Los archivos utilizados aparecen en la **Tabla 18**.

Identificador	Nombre	Descripción	Formato
DMG05	DMG05_DeliRCyFMariaGrandeE10000.wmf	Plano de la planta urbana de María Grande donde están demarcados radios censales y fracciones.	WMF
DC02	DC02_1P-FR.xls	Total de población según sexo por departamento, por área de gobierno local, por fracción y radio censal.	XLS
DC03	DC03_6H-FR.xls	Hogares habitados según procedencia del agua para beber y cocinar por departamento, área de gobierno local, fracción y radio censal.	XLS

Tabla 18: Archivos consultados del Censo 2010 para la determinación del ratio habitantes por vivienda y radio censal.

3.4.1.2.1.2. Depuración de los registros de consumo y tratamiento con SIG de la información disponible

En primer lugar, se escanea el registro DMG01. A través de la aplicación de un programa con reconocimiento óptico de caracteres (proceso conocido por sus siglas en inglés OCR) se transforman las imágenes obtenidas a MS Word. Seguidamente, las tablas identificadas se copian en una hoja de cálculo y se lleva a cabo una inspección visual para corregir de forma manual los errores cometidos por el OCR. El resultado se guarda en *DMG01_ConsumoAnualMedioAsoCaracParcMariaGrande.xlsx*.

En segundo lugar, las lecturas bimensuales de contadores procedentes de María Grande (archivos DMG03 y DMG04) se someten a una serie de filtros, con el objetivo de excluir aquellos registros que pudieran contener errores. Para ello se conforma una hoja de cálculo, (*R1_RegistroConsumosMensuales20132014MG.xlsx*) en la que, previa determinación de los volúmenes consumidos en cada bimestre, se introducen las siguientes condiciones a cumplir por un registro para que éste pueda ser considerado como fiable:

- i. $L_{t+1} > L_t$: La lectura realizada en el mes t+1 tiene que ser mayor que la lectura en el mes t.
- ii. $L_t <> ""$ (vacío): No deben haber saltos en las lecturas, es decir, no se permiten registros sin dato en algún mes.
- iii. $(V_t - V_m) / V_m > 1.5$: La desviación entre el volumen consumido en un bimestre y el volumen medio de todos los bimestres disponibles, no puede superar 1.5 veces dicho volumen medio.

De 1984 registros, son descartados 700 después de aplicar los filtros. El volumen bimestral de interés es el que corresponde a los meses de noviembre y diciembre de 2013, el cual se calcula como la resta entre la lectura 1 del registro DMG04 (realizada entre noviembre y diciembre de 2013) y la lectura 6 del registro DMG03 (realizada entre septiembre y noviembre de 2013), ya que en Nogoyá las mediciones de caudal inyectado desde el tanque principal nro. 35 se realizaron en diciembre.

A continuación, se procesa la información de DMG02, DMG05, DC02 y DC03 con ArcGIS, creando un bloque análogo a la unión de los BLOQUES 1 (Trama urbana) y 4 (Censo 2010) del proyecto SIG iniciado en Nogoyá, pero sin llevar a cabo la georeferenciación de la misma por falta de datos. Las operaciones realizadas en ArcGIS para alcanzar los resultados finales mostrados en la **Tabla 19** son iguales a las expuestas en los apartados 3.4.1.1.2 y 3.4.1.1.5.

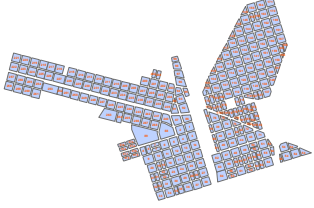
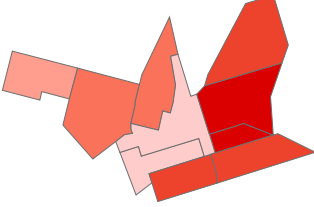

Nombre de la capa	Descripción	
105_104_102_Manz	Capa de polígonos que incluye las manzanas y su correspondiente identificador.	
111_107_Join_Part	Capa de polígonos en la que están delimitados los radios censales. Incluye información sobre vivienda y población totales por radio censal ⁸ , así como ratio de habitantes por vivienda promedio.	
113_105_112_SpatialJoin_PartidaAutocadCompleta	Capa de puntos en la que cada uno de ellos representa una parcela dibujada en DMG02 y contiene su correspondiente identificador. En el caso de que éste coincidiera a la vez con los incluidos en los registros DMG01 y DMG03-DMG04, se le ha asociado a la parcela correspondiente la información que la caracteriza, su consumo medio anual en 2013 y el consumo medio mensual en el bimestre noviembre-diciembre de 2013. Por último, mediante unión espacial se han añadido los campos de nro. de manzana y habitantes por vivienda promedio, procedentes de las anteriores capas resultado.	

Tabla 19: Resultados alcanzados al introducir en ArcGIS la información de DMG02, DMG05, DC02 y DC03.

Por último, se uniforma el identificador de parcela, llamado *Partida*, y se analiza la concordancia del mismo entre los distintos registros, con el propósito de cruzar la información que recogen DMG03, DMG04 y la capa procesada *113_105_112_SpatialJoin_PartidaAutocadCompleta*. Para ello se conforman las hojas de cálculo *R1_CorrespID_AutoCAD_RegistroPapel.xlsx* y *R1_RegistroConsumosMensuales-20132014MG.xlsx*.

Como resultado final de este sub-apartado se obtiene una tabla con 989 registros, que sirve como dato de entrada al proceso de clústering. Cada uno de ellos corresponde a una parcela de María Grande, que está caracterizada por su superficie total, superficie construida y ratio habitantes por vivienda promedio.

⁸ Unidad espacial mínima en los estudios realizados por la Dirección de Estadísticas y Censos de la Provincia de Entre Ríos.

3.4.1.2.1.3. Análisis estadístico univariado y bivariado del registro de parcelas depurado.

El presente sub-apartado tiene como objetivo conocer a nivel de estadística descriptiva cómo es cada una de las variables que intervendrán de un modo u otro en el clústering de parcelas de María Grande.

Primeramente, se realiza un análisis estadístico univariado. Los resultados obtenidos se resumen en las Tablas 20, 21, 22 y 23.

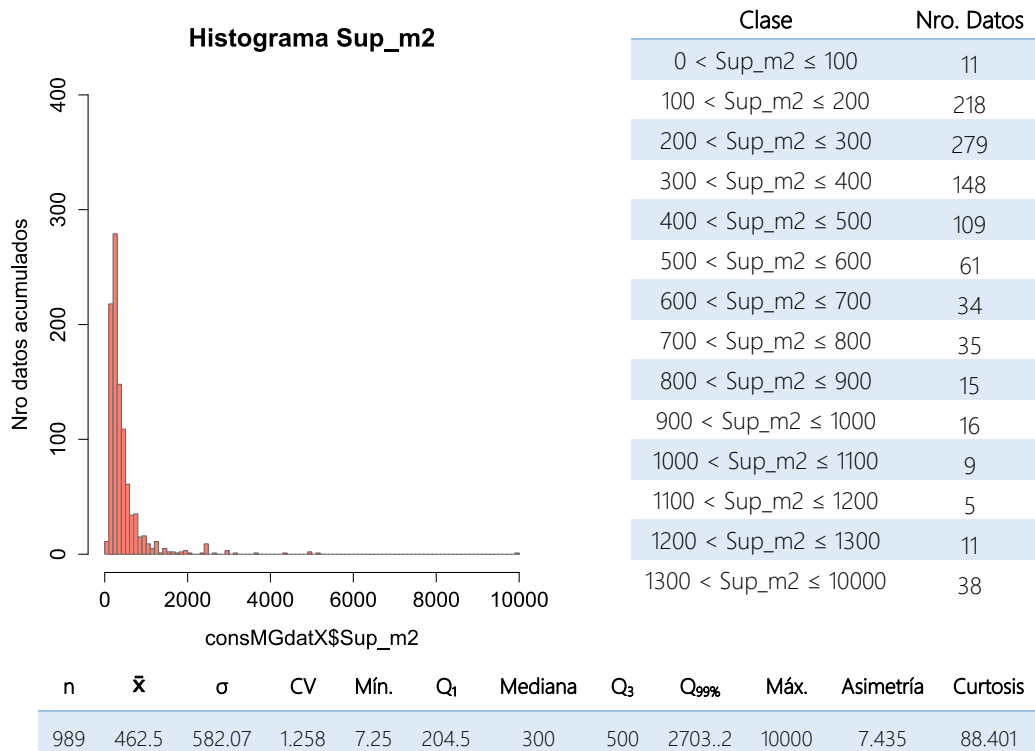


Tabla 20: Resumen del análisis univariado de la *superficie total de la parcela* (Sup_m2).

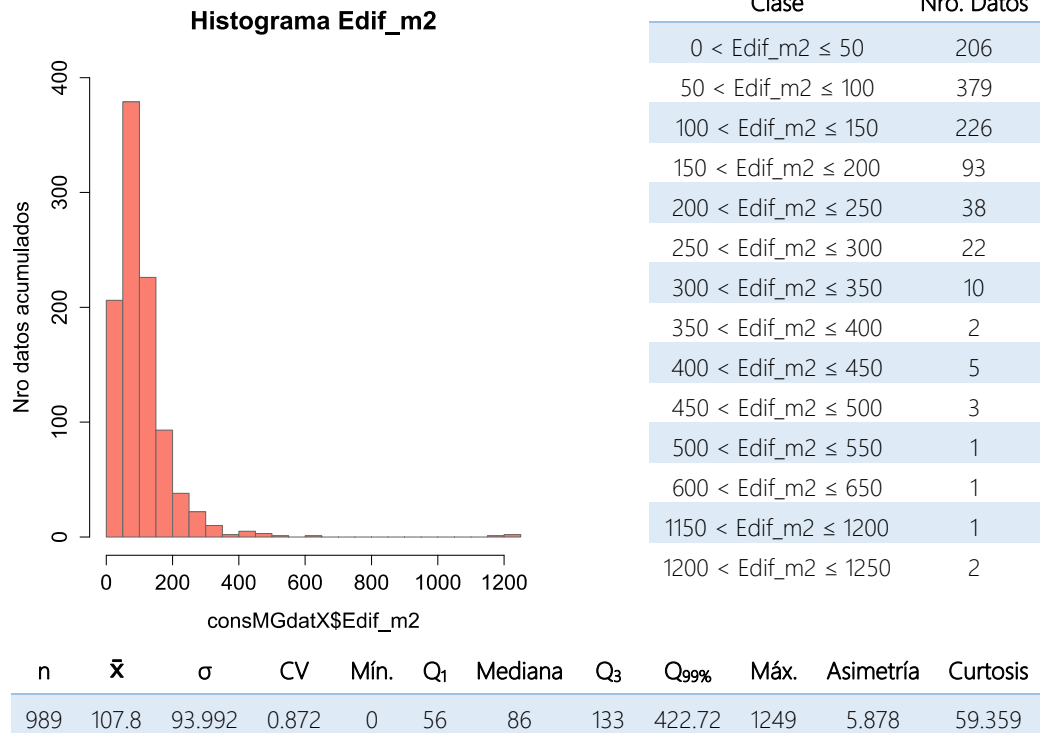


Tabla 21: Resumen del análisis univariado de la *superficie construida* (Edif_m2).

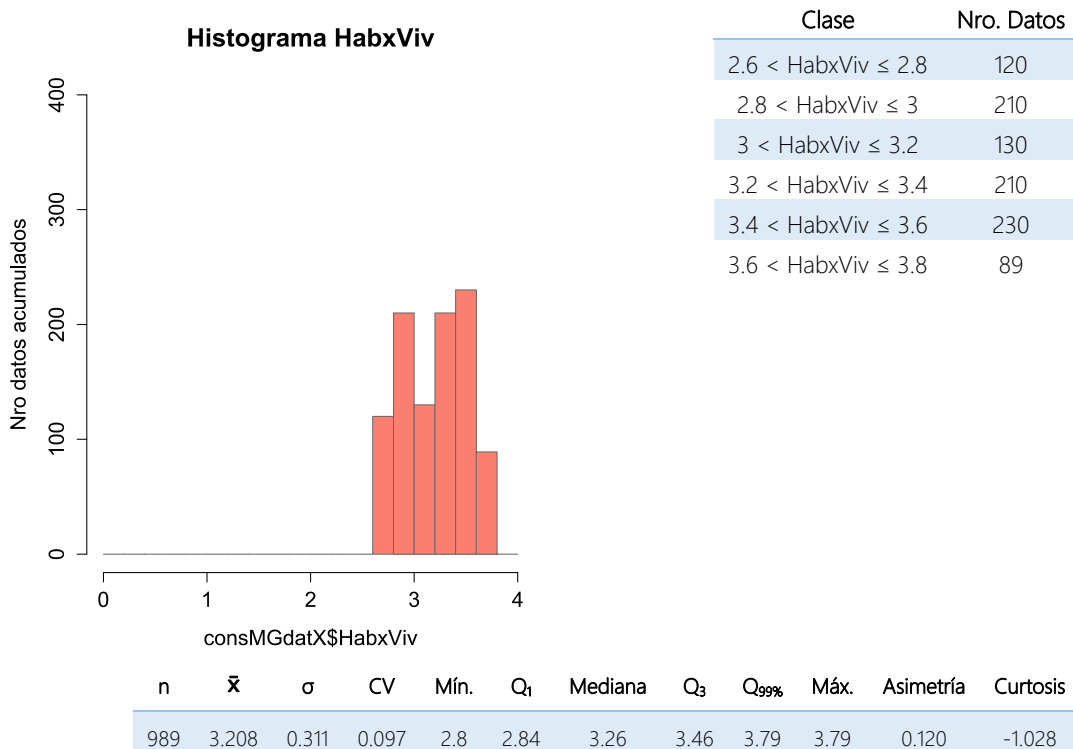


Tabla 22: Resumen del análisis univariado de los *habitantes por vivienda* (HabxViv).

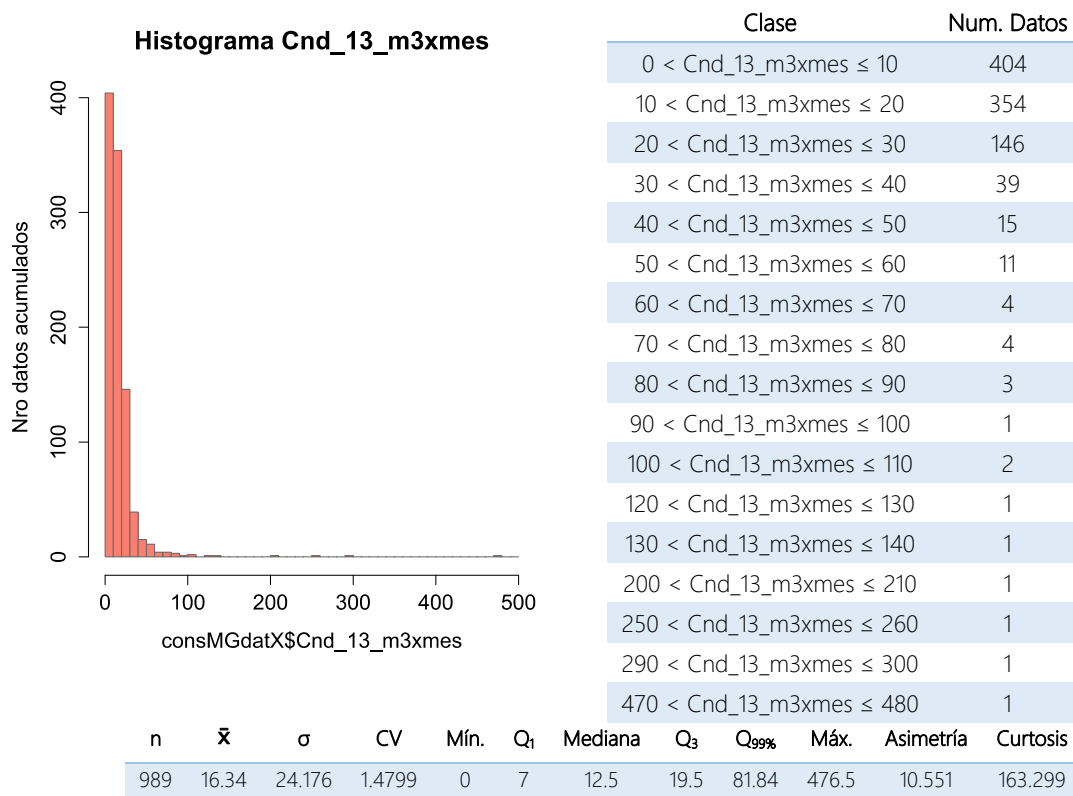


Tabla 23: Resumen del análisis univariado del consumo medio en el bimestre noviembre-diciembre (Cnd_13_m3xmes).

Las conclusiones más importantes que se extraen son:

- La variable *habitantes por vivienda* se ha obtenido a partir de los datos del Censo 2010. La unidad espacial mínima a la que hace referencia dicho estudio es el *radio censal*, zona urbana que abarca varias manzanas. El cambio de escala asociado con respecto a la entidad parcela, provoca que su comportamiento diste notablemente del resto de variables.
- Tanto la variable *superficie total de la parcela* como la variable *superficie construida* presentan asimetría positiva ($\text{Mediana} < \bar{x}$ y el coeficiente de asimetría superior a cero) y una forma leptocúrtica, es decir, existe una mayor concentración de datos cerca de la media, coexistiendo con una frecuencia no despreciable de datos alejados de la misma. Además, en ambos casos el valor máximo es notablemente superior a $Q_{99\%}$, que corresponde al valor de la variable para una frecuencia acumulada del 99%. Se ha comprobado que dichos extremos no son errores de transcripción o valores atípicos respecto a los que se dan en la población de Nogoyá.
- La variable *consumo medio en el bimestre noviembre-diciembre* es también asimétrica y leptocúrtica, así como $Q_{99\%}$ dista notablemente del valor máximo. Cabe destacar que entre estas dos medidas de posición se encuentran los tres consumos más importantes en orden de magnitud (asociados a las tres parcelas que poseen

mayor superficie construida), que corresponden al Hospital Francisco Casals, el Club Deportivo de María Grande y un colegio (a partir de ahora, el acrónimo *CI* ("consumos importantes") hará referencia a estos tres individuos de la muestra).

A continuación, se lleva a cabo un análisis estadístico bivariado con el propósito de observar qué grado de correlación guarda el consumo con las variables consideradas para el clústering de parcelas. El Coeficiente de Correlación de Pearson, calculado teniendo en cuenta el grupo *CI* y sin él, se resume en la **Tabla 24**. En la **Figura 15** se puede consultar los gráficos de dispersión propios de este análisis.

<i>Matriz de correlaciones incluyendo el grupo CI</i>	<i>Matriz de correlaciones sin incluir el grupo CI</i>
Sup_m2 Edif_m2 HabxViv	Sup_m2 Edif_m2 HabxViv
Cnd_13_m3xmes 0.250 0.584 -0.062	Cnd_13_m3xmes 0.027 0.209 -0.062

Tabla 24: Coeficiente de Correlación de Pearson entre el consumo y las variables consideradas para el clústering de parcelas.

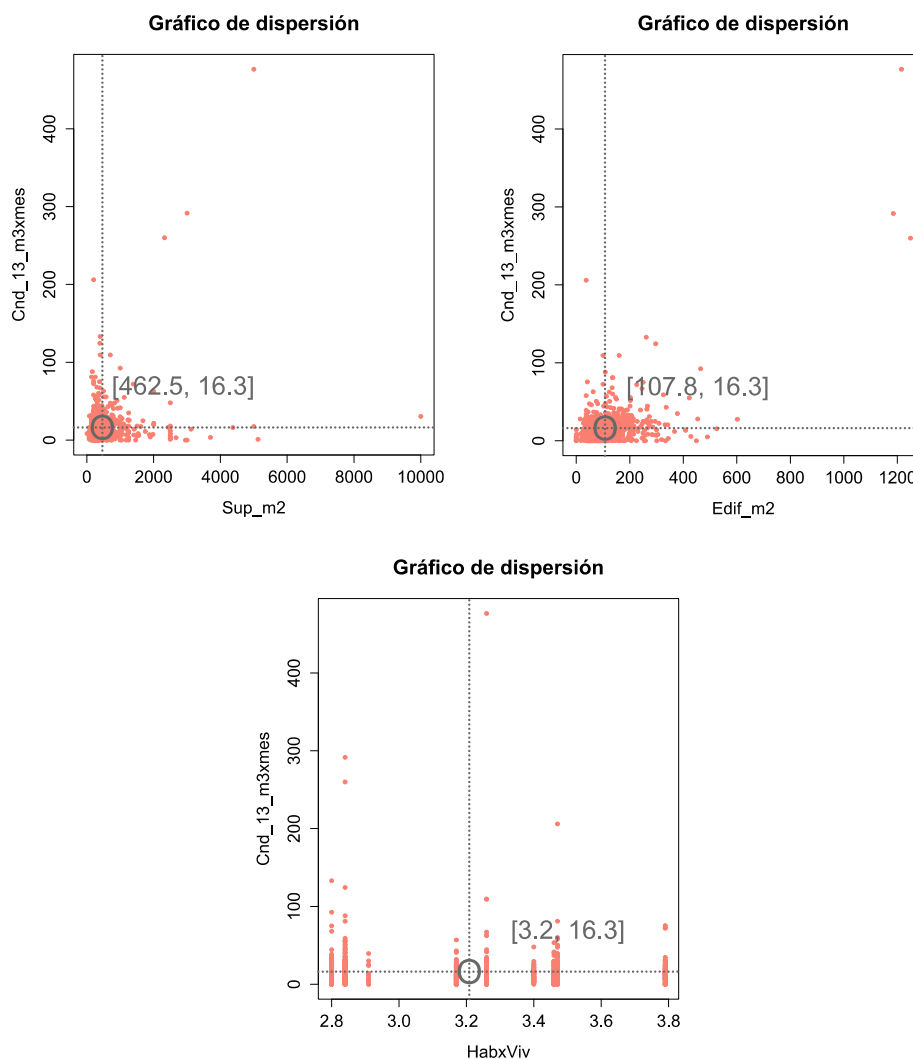


Figura 15: Gráficos de dispersión que enfrentan las variables implicadas en el clústering de parcelas con su correspondiente consumo medio en los meses de noviembre y diciembre, *CI* incluidos.

Las observaciones más destacables son:

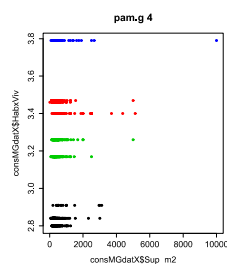
- El *consumo medio de noviembre-diciembre* guarda correlación positiva de forma tenue con la *superficie total* de la parcela y la *superficie construida*. No obstante, no hay que perder de vista que el Coeficiente de Correlación de Pearson está fuertemente afectado por los puntos que quedan lejos de la nube principal. En este caso, los asociados al hospital, el colegio y el club deportivo afectan determinantemente al valor de la correlación a la que se ha hecho referencia, de ahí que al no tenerlos en cuenta se anule prácticamente la misma.
- La variable *habitantes por vivienda* no está correlacionada con el *consumo medio de noviembre-diciembre*. Se cree que la principal razón es que está definida a escala de radio censal.

3.4.1.2.1.4. Justificación del descarte de la variable habitantes por vivienda promedio para el proceso de clústering de parcelas.

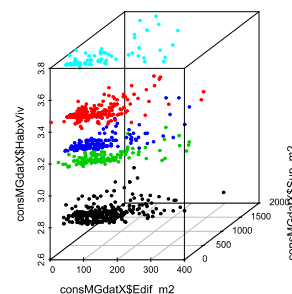
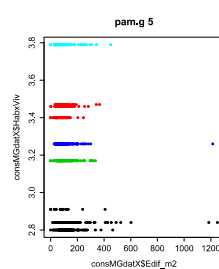
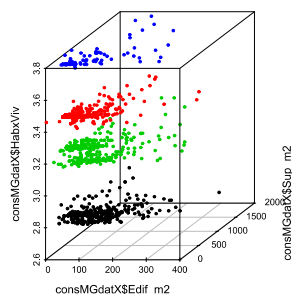
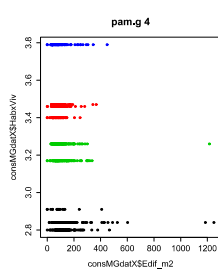
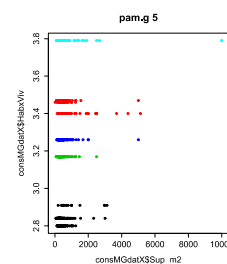
Los primeros análisis de clústering se realizan con las tres variables determinadas para caracterizar una parcela: la *superficie total*, la *superficie construida* y los *habitantes por vivienda*. Todas ellas son definidas como variables numéricas.

Las agrupaciones resultado obtenidas conducen a la siguiente conclusión: la variable *habitantes por vivienda* provoca una categorización de los clústeres en función de su valor, tal y como puede observarse en la **Figura 16**, que incluye gráficos 3D de una de algunas de las soluciones. De ello se infiere que incluirla genera agrupaciones sesgadas y que, por tanto, no satisfacen el objetivo a alcanzar. Se cree que esto es debido principalmente al cambio de escala que requiere el cálculo de esta variable con respecto a la entidad de referencia, la parcela. Una posible solución sería tratar esta variable como categórica, pero para este trabajo se ha decidido suprimirla.

PAM. 4 clústeres.
Métrica Gower.



PAM. 5 clústeres.
Métrica Gower.



PAM. 10 clústeres.
Métrica Gower.

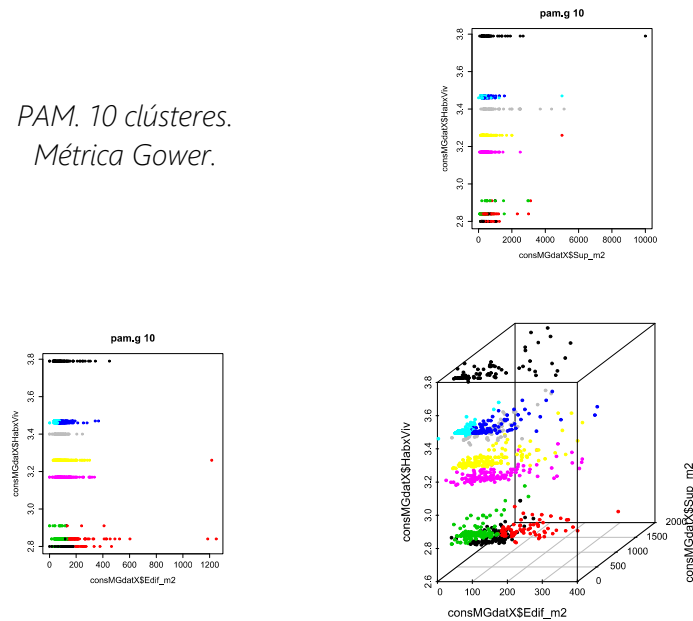


Figura 16: Gráficos 3D que muestran las soluciones del clústering con el algoritmo PAM y variables de entrada superficie total de la parcela, superficie construida y habitantes por vivienda.

3.4.1.2.1.5. Clústering de parcelas

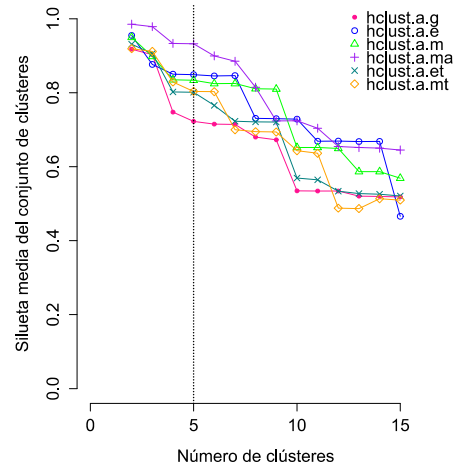
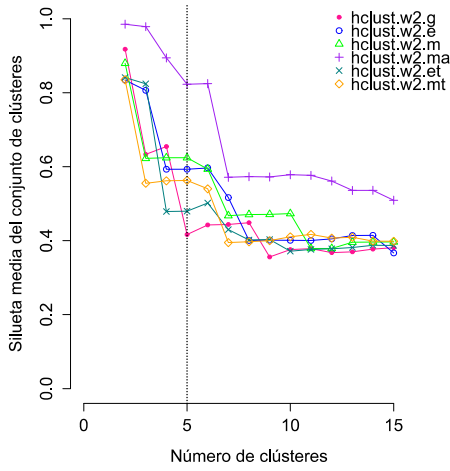
Los tipos de clústering realizados para la agrupación de parcelas de María Grande son:

- Clústering jerárquico aglomerativo
 - Método ward.D2
 - Método UPGMA
- Clústering jerárquico divisivo DIANA
- Clústering de partición
 - KMEANS
 - PAM

Los resultados arrojados por todos ellos se analizaron globalmente a través de la visualización de los gráficos que aparecen en la **Figura 17**, en los cuales se representa el valor de silueta media del conjunto de grupos por número de clústeres y para cada métrica, en su caso.

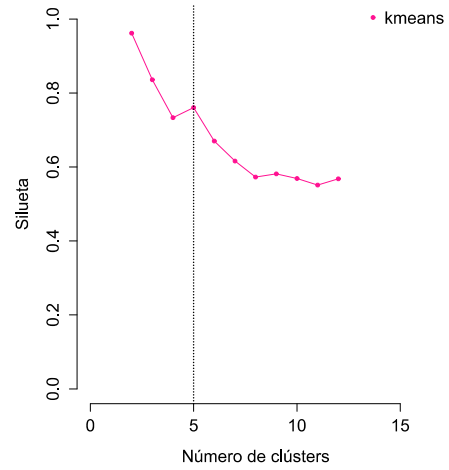
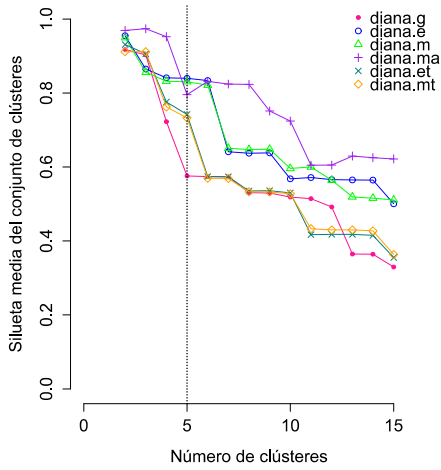
Clústering jerárquico aglomerativo. Método ward.D2.

Clústering jerárquico aglomerativo. Método UPGMA.



Clústering jerárquico divisivo DIANA

Clústering de partición KMEANS



Clústering de partición PAM

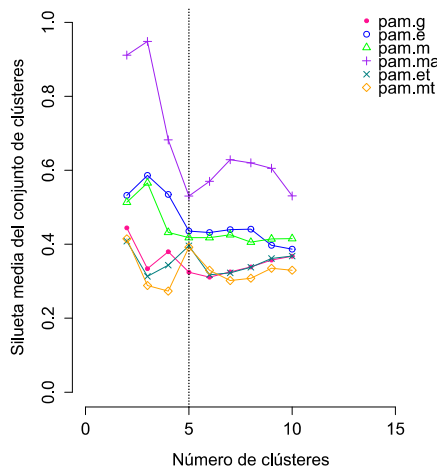


Figura 17: Gráficos asociados a cada tipo de clústering empleado que contrapone el valor de silueta media del conjunto de clústeres con el número formado de estos para cada métrica utilizada, en su caso.

La percepción en la mayor parte de los casos es que 5 sería el número máximo de clústeres admisibles, ya que a partir de dicho número el valor de silueta del conjunto de grupos empieza a decaer notablemente. Es decir, un aumento del número de agrupaciones podría estar generando clústeres artificiales, dado que la proximidad entre algunos de ellos sería mayor que la similitud de los individuos que los forman. Esta es la razón por la cual 5 es el número de grupos seleccionado para el resultado del presente clústering.

El siguiente paso fue editar, para cada uno de los procesos de clústering nombrados, 2 gráficos asociados a las soluciones que contemplan 5 clústeres:

- Un gráfico de dispersión que enfrenta las dos variables de entrada (Sup_m2 y Edif_m2), donde pueden distinguirse los grupos conformados según color.
- La silueta resultado, que incluye el tamaño de cada grupo, la silueta media de grupo y la silueta media del conjunto de grupos.

Los mismos pueden ser consultados en el Anejo 1. De su análisis se deduce:

- La distancia Euclidiana y Manhattan han demostrado no ser una buena opción, ya que están afectadas de forma notable por el orden de magnitud de cada una de las variables, tal y como evidencia que la distinción de grupos esté determinada por la superficie total de la parcela. Es por ello que quedan descartadas todas las soluciones que emplean estas métricas como base, incluido el clústering de partición KMEANS.
- El método UPGMA (clústering jerárquico aglomerativo) y el algoritmo DIANA (clústering jerárquico divisivo), presentan el problema de dar mayor peso en la conformación de clústeres a aquellos grupos que son más numerosos. Debido a esto, no aportan soluciones interesantes para los objetivos perseguidos.
- Las soluciones asociadas a las métricas que no se ven afectadas por el cambio de escala de la variable, como es el caso de la distancia Gower o Mahalanobis, así como aquellas resultado de aplicar la distancia Euclidiana y Manhattan a la variable tipificada, resultan interesantes y requieren atención:
 - La solución proporcionada por el método ward.D2 (clústering jerárquico aglomerativo) y métrica Mahalanobis, obtiene valores de silueta media del conjunto de clústeres mayores que el resto de soluciones. Desde el punto de vista estricto, es el clústering que diferencia agrupaciones más compactas y, en este sentido, mejor conformadas. Además, define un clúster específico para los *CI*. Si bien, la solución proporcionada no genera grupos dentro de la nube principal de puntos, donde se encuentran casi el 90% de los individuos de la muestra analizada.
 - Las soluciones que proporcionan el método ward.D2 (clústering jerárquico aglomerativo) y PAM (clústering de partición) para la métrica Gower, Euclidiana y Manhattan, aplicadas las dos últimas a las variables de entrada

tipificadas, son muy parecidas. Esto es un indicador de que esa estructura dentro de los datos es plausible. Además, sí genera una subdivisión dentro de la nube principal de puntos, rasgo que resulta interesante dado que ese era el principal objetivo del presente clústering. No obstante, los valores de silueta alcanzados son peores que la solución anteriormente nombrada y los *CI* no quedan claramente diferenciados.

Finalmente, en la elección de la solución adoptada ha pesado más el generar varios grupos con un tamaño importante que la compacidad de los mismos, representada por la silueta. Se cree que el resultado generado por el clústering *de partición PAM con métrica Euclidiana aplicada a las variables tipificadas*, es el que cumple este criterio de forma más óptima. Si bien, se ha forzado a que los *CI* queden dentro de un sexto grupo, ya que claramente constituyen un clúster distinto desde el punto de vista del consumo y del uso al que se destinan esas parcelas.

El gráfico de dispersión de la solución adoptada aparece en la **Figura 18**.

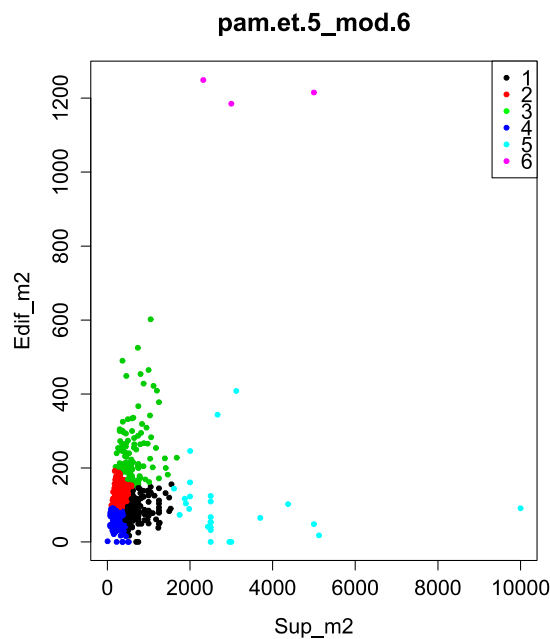


Figura 18: Solución adoptada para el clústering de parcelas de María Grande.

Este resultado constituye el dato de entrada para el entrenamiento de un SVM, herramienta que permite emular la clasificación de parcelas en María Grande y, por tanto, reproducir la misma en Nogoyá. La formulación de SVM escogida es la conocida como C-Support Vector Classification (Chih-Chung, 2001), en la que se han utilizado las opciones por defecto en lo que a la función kernel se refiere, es decir, función de base radial tipo gaussiana con parámetro $\gamma=1/[\text{nro. componentes de los vectores de entrenamiento}]$.

Como técnica de validación se aplica el Método de Retención (*Holdout Method*), así que los 989 vectores de 2 dimensiones (superficie construida y superficie total) que forman el registro de parcelas depurado de María Grande se dividen en 3 grupos distintos: entrenamiento, validación y test. A cada uno de ellos le corresponde un porcentaje de los datos del 60% (600), 20% (195) y 20% (194), respectivamente. A su vez, el grupo de validación se sub-divide en 13 grupos de 15 vectores, de forma que el desempeño del SVM se evalúa a través de la media de los errores cometidos en cada uno de ellos. Añadir que las variables de los vectores de entrenamiento fueron tipificadas, ya que de este modo la clasificación resultaba más satisfactoria.

Los grupos de entrenamiento y validación fueron empleados para encontrar el valor óptimo del parámetro **C**. Los valores candidatos están comprendidos entre 1 y 1500, separados por intervalos de 50 unidades. Para cada uno de dichos valores se llevan a cabo los siguientes pasos:

1. Se llama a la función *svmtrain*, de la librería LIBSVM (adaptada a Matlab), y se le pasa como parámetros el set de vectores de entrenamiento $\mathbf{x}_i \in \mathbb{R}^2$, $i = 1, \dots, 600$, su correspondiente vector $\mathbf{y} \in \mathbb{R}^{600}$, tal que $y_i \in \{1, 2, 3, 4, 5, 6\}$, con la clasificación, y el parámetro **C** candidato.
2. Se llama a la función *svmpredict* para que clasifique los 15 vectores de cada uno de los grupos de validación con el SVM entrenado en el anterior paso. El error para cada grupo se calcula como: $Error_{grupo\ j} = \sum_{i=1}^{15} (y_i^{clas_predicha} - y_i^{clas_dato})$, $j = 1, \dots, 13$.
3. Dado que el promedio de los $Error_{grupo\ j}$ representa el desempeño del parámetro **C** candidato, se almacenan ambos en la misma fila de una tabla.

El **C** con menor promedio de los $Error_{grupo\ j}$ es el escogido como óptimo.

Se comprueba la precisión del SVM calibrado con el grupo de test: el 97.2 % de los vectores incluidos en él son correctamente clasificados. En consecuencia, se puede afirmar que la calibración del SVM es válida y está listo para ser aplicado a los datos de Nogoyá.

Para cualquier aclaración conceptual acerca del clústering de parcelas de María Grande, véase el apartado 2.1.

3.4.1.2.2. Clasificación de las parcelas de Nogoyá

Llegados a este punto, se hace necesario obtener y editar la información considerada para caracterizar una parcela, pero en esta ocasión sita en Nogoyá.

La superficie total y construida se extrajo vía web del visor SIG habilitado por la empresa GEOMAT, encargada de mantener actualizado el Catastro del municipio. Se invirtieron

esfuerzos en intentar conseguir la información sobre Nogoyá que esta empresa disponía en formato capa SIG, con la que se podría haber ahorrado el trabajo expuesto en el apartado 3.4.1.1, pero finalmente no fue posible. La descarga de datos sobre las parcelas de Nogoyá desde el visor SIG es extremadamente ineficiente, ya que se realiza en formato PDF y en una consulta sólo permite seleccionar un número limitado de parcelas. Los pasos seguidos para adquirir estos datos son:

1. Seleccionar la parcelas contenidas en las manzanas delimitadas en la capa *105_104_708_SpatialJoinEditTableText_Man*. Para asegurar que la información de todas ellas era descargada, la zona de selección en cada consulta quedaba circunscrita a una o dos manzanas. Era inevitable que dichas zonas de selección se solaparan (Figura 19).

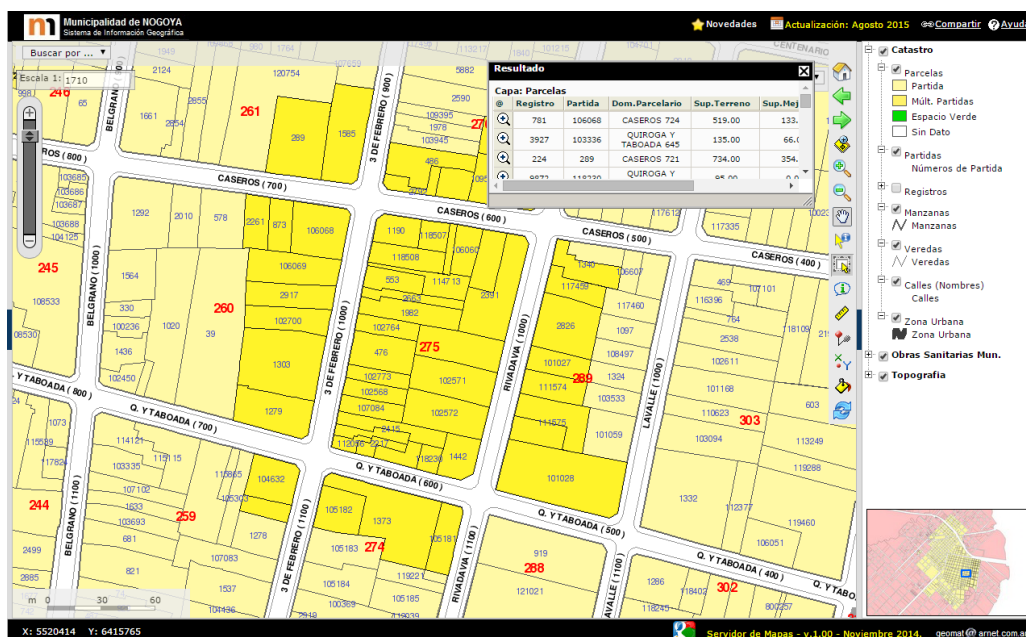


Figura 19: Captura de pantalla de la web de GEOMAT en la que puede observarse la selección de parcelas.

2. Descargar en PDF la información asociada (Figura 20).

Parcelas

Registro	Partida	Dom.Parcelario	Sup.Terreno	Sup.Mejoras	Manzana	Fecha.Dato
1635	2068	SAN MARTIN 236	431.00	81.00	237	20150309
21036	121509	SAN MARTIN 238	251.00	89.00	237	20150309
6293	109085	SAN LORENZO	2456.00	0.00	1523	20150309
3278	101261	BV.RAMIREZ	2098.00	97.00	1523	20150309
4688	105694	RUJA NACIONAL Nº12	338.00	87.00	1523	20150309
5928	108398	RUJA NACIONAL 12	269.00	30.00	1523	20150309
6115	108677	9 DE JULIO	728.00	126.00	205	20150309
20935	116097		835.00	130.00	205	20150309
3465	101854	9 DE JULIO	835.00	255.00	205	20150309
4775	105892	9 DE JULIO Nº 50	268.00	127.00	221	20150309

Figura 20: Extracto de uno de los PDF descargados del visor SIG habilitado por GEOMAT.

- Los PDF se abren con MS Word. Las tablas resultado son copiadas a una hoja de cálculo. En la misma se escribe una macro que suprime aquellos registros duplicados debido a los solapes de las zonas de selección.

Cabe destacar que la fecha de actualización de los datos empleados es el 9 de marzo de 2015. En consecuencia, posteriores modificaciones del Catastro no han sido tenidas en cuenta.

Al intentar asociar la anterior información a la capa de parcelas *205_204_710_SpatialJoin_Parc*, queda patente que el AutoCAD *DTU03* dato está desactualizado con respecto a la delimitación e identificación de parcelas. No obstante, es posible ubicar en el espacio dicha información gracias a la columna en la que está definido el número de manzana a la que pertenece cada parcela. Para ello, hay que cruzar la información procedente del visor SIG y del AutoCAD *DTU03*, almacenada en la capa *105_104_708_SpatialJoinEditTableText_Manz*, y, por tanto, uniformar los identificadores de manzana en ambos registros. Esta última operación requiere modificaciones en la geometría de algunos polígonos de *105_104_708_SpatialJoinEditTableText_Manz*, así como decidir cuál va a ser el identificador de referencia y realizar las correspondencias pertinentes con los identificadores originales. También hay que añadir la depuración de errores en los datos obtenidos del visor SIG. Este trabajo queda inscrito en la hoja de cálculo *R2_ParcSinDuplicadosCorrespIDManz.xlsm* y el resultado final plasmado en la capa SIG *901_105_EditTableIDUni_DemandManz*.

A continuación, sólo resta preparar una tabla con la superficie total y construida de cada parcela, que servirá de dato de entrada al SVM entrenado y validado. La clasificación que da como resultado el mismo está graficada en la **Figura 21**.

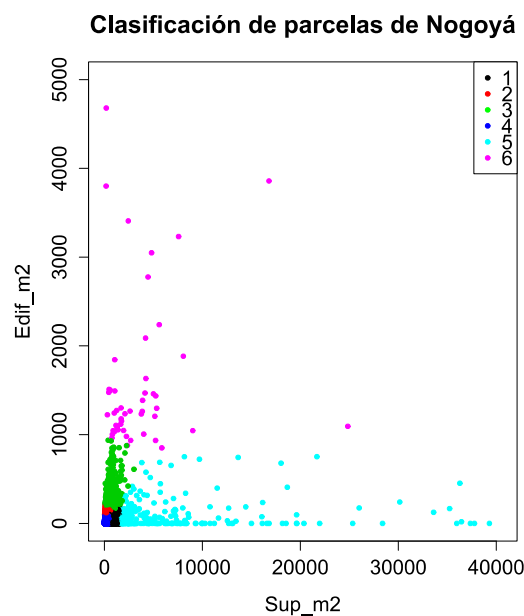


Figura 21: Clasificación de las parcelas de Nogoyá proporcionada por el SVM entrenado y validado.

3.4.2. Tratamiento de información que interviene en la fase de calibración

3.4.2.1. Caudales

El registro de mediciones tomadas con el caudalímetro descrito en el apartado 3.3.2.1 se post-procesa a través de la hoja de cálculo *R3_DatosCaudalimetro.xlsm*. Éste consta de los siguientes campos:

- Volumen acumulado por el caudalímetro, instalado en la tubería de subida al tanque principal nro. 35, en m^3 .
- Presión aguas abajo de la VRP ubicada en cabecera de red, en mca.
- Nivel del tanque principal nro. 35, en mca.

Se recuerda que los datos se recababan de forma manual al inicio de cada hora y que el registro abarca el periodo comprendido entre el día 2 a las 00:00 horas y el día 13 a las 00:00 horas de diciembre de 2014.

Las operaciones realizadas son:

1. Organizar los registros en una única tabla, indexada según la posición de la hora de la semana (que toma el valor 0 para el lunes a las 00:00 y el valor 167 para el domingo a las 23:00).
2. Calcular el caudal de salida del tanque, en m^3/h , aplicando la ecuación de continuidad al mismo.
3. Para los días de la semana repetidos, se calcula la media del caudal inyectado, tal y como se muestra en la **Figura 22**.

Añadir, que existen tres errores de medición del caudal, sobrevenidos el jueves 4 de diciembre de 7:00 a 8:00, el sábado 6 a las 19:00 y el viernes 12 a la 1:00.

Finalmente, se decide que la curva representativa del caudal inyectado desde el tanque principal nro. 35 es la que se obtiene como la media entre las dos semanas de medición. Los errores que tuvieron lugar el jueves y viernes se pudieron sortear con los datos correspondientes de la semana sin error. En cambio, el fallo en la medición del sábado a las 19:00 tuvo que suplirse realizando la media entre el valor precedente y subsiguiente.

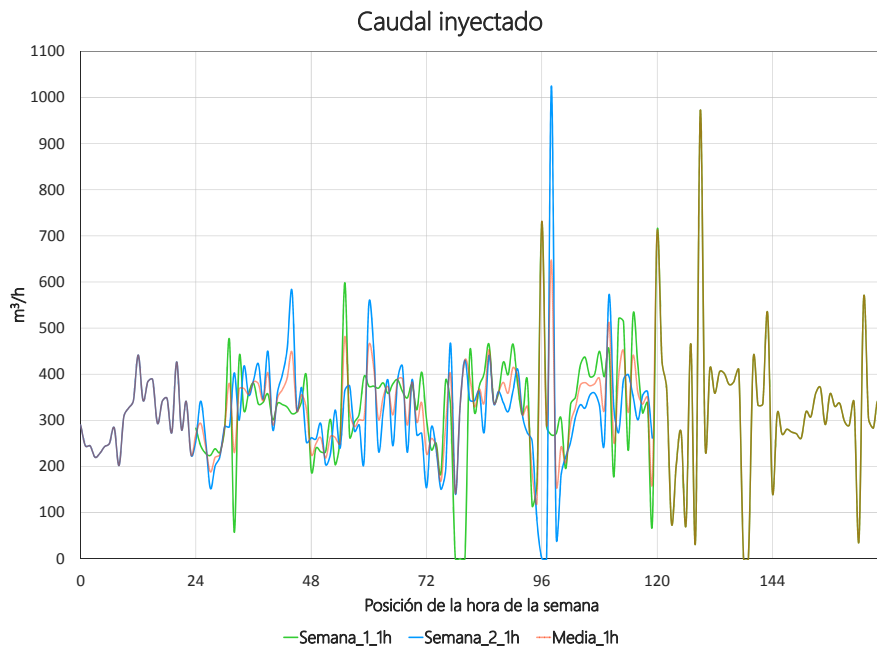


Figura 22: Gráfico donde aparece el caudal inyectado a red la semana 1 (del 2 al 7 de diciembre de 2014) y 2 (del 8 al 12 de diciembre de 2014) de mediciones.

Para terminar, con el propósito de suavizar la curva seleccionada y corregir oscilaciones de la medición asociadas al ruido, el cual puede deberse a diversas causas (por ejemplo, efectos transitorios causados por el arranque o parada de una electrobomba), se calcula la media móvil simple considerando un dato previo. La curva del caudal inyectado resultado se puede observar en la **Figura 23**.

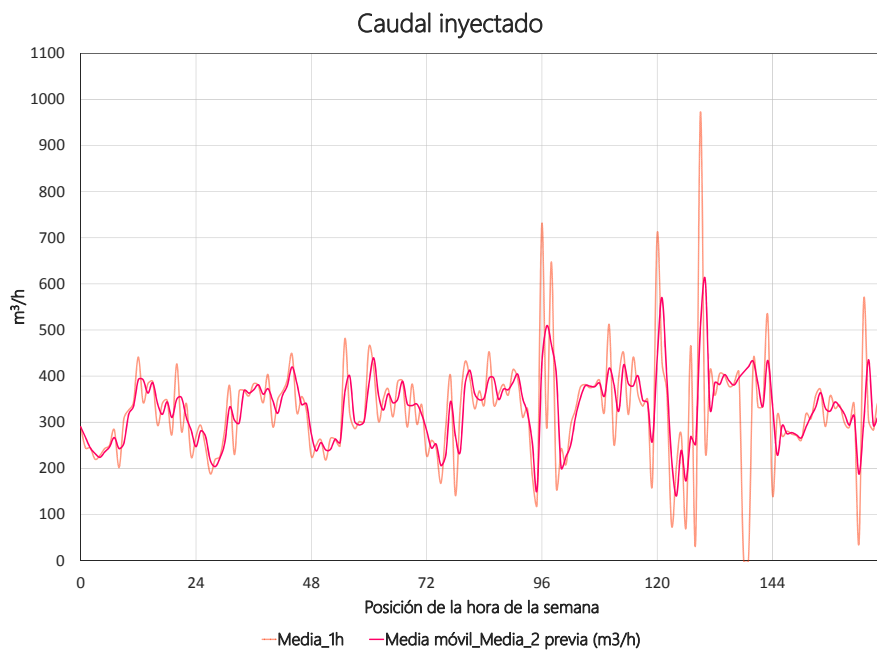


Figura 23: Gráfico en el que aparece la curva de caudal inyectado empleada como dato de calibración, resultado de aplicar la media móvil simple con un dato previo a la media de las mediciones tomadas en las semana 1 y 2.

3.4.2.2. Presiones

En el gráfico de la **Figura 24** se halla la curva resultado de un proceso análogo al explicado anteriormente, pero en esta ocasión aplicado a los datos de presión medida aguas abajo de la VRP ubicada en cabecera de red. Es conveniente aclarar que al valor registrado se le ha sumado 1.5 metros, para tener en cuenta la profundidad a la que está enterrada la tubería y la altura sobre el terreno a la que se encuentra el medidor de presión, así como la cota del terreno, ya que dicha curva se va a emplear en la modelación del punto de inyección que constituye el tanque principal nro. 35.

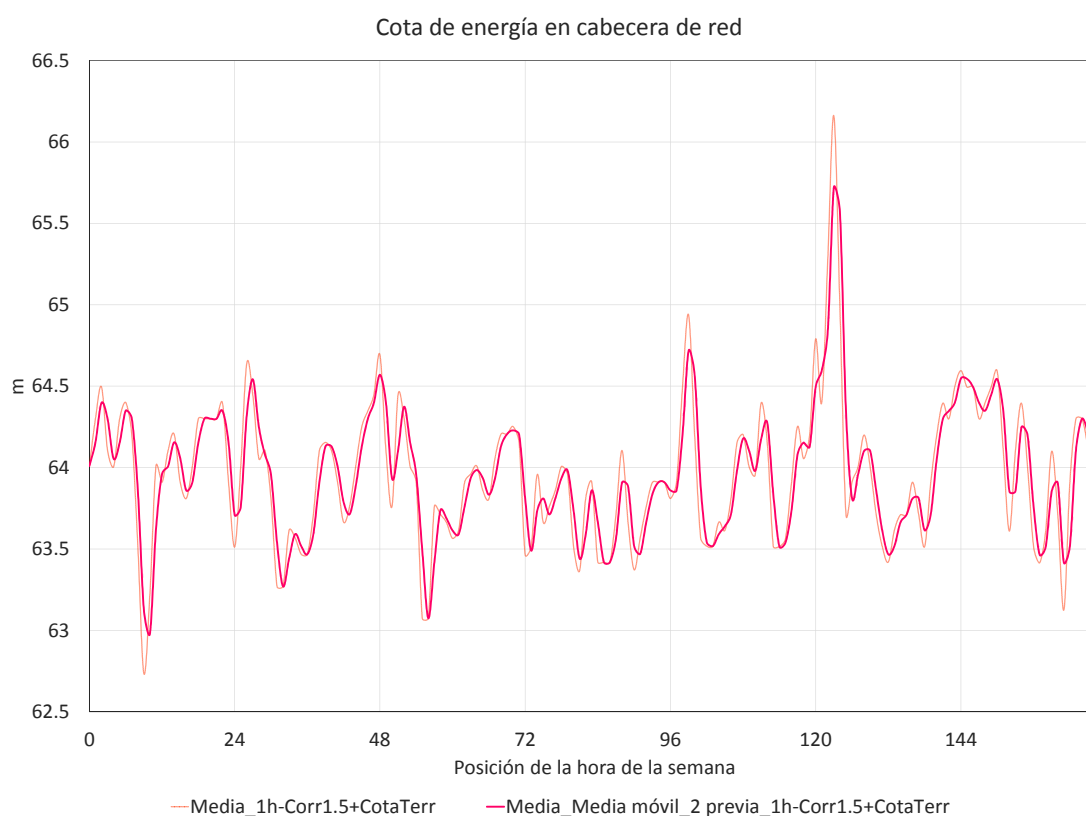


Figura 24: Gráfico en el que aparece la curva que describe la cota de energía aguas abajo de la VRP.

Los datos de las presiones observadas en puntos distribuidos por la red se han tratado con la asistencia de una hoja de cálculo (*R4_Mediciones_Presion_Verano.xlsm*) del siguiente modo:

- En tanto que se encuentran almacenados en archivos planos, se copian a la hoja de cálculo y estructuran para su análisis.
- Cada uno de los registros es distribuido temporalmente a escala semanal. También se identifica la época del año a la que pertenecen (la época primavera-verano corresponde a los meses comprendidos entre octubre y marzo, inclusive, y la época otoño-invierno a los meses entre abril y septiembre) y se corrige sus valores

sumándoles 1.5 mca, para tener en cuenta la profundidad a la que está enterrada la tubería. La **Tabla 25** resume el resultado de este paso.

ID	LUNES	MARTES	MIÉRCOLES	JUEVES	VIERNES	SÁBADO	DOMINGO
P1-1				06/12/2012 12:00		08/12/2012 12:00	
P1-2				26/07/2012 18:00		28/07/2012 18:20	
P2-1		30/10/2012 12:00			02/11/2012 12:00		
P2-2		12/03/2013 12:00		14/03/2013 12:00			
P3-1			12/12/2012 12:00		14/12/2012 12:00		
P3-2		14/08/2012 12:00			10/08/2012 12:00		
P4-1	15/10/2012 12:00				12/10/2012 12:00		
P5-1		18/09/2012 12:00			21/09/2012 12:00		
P5-2			19/12/2012 12:00		21/12/2012 12:00		
P6-1		18/09/2012 12:00			21/09/2012 12:00		
P6-2			16/01/2013 12:00		18/01/2013 12:00		
P6-3			30/01/2013 12:00		01/02/2013 12:00		
P7-1		11/09/2012 12:00			14/09/2012 12:00		
P8-1			29/08/2012 11:00			01/09/2012 11:00	
P8-2			09/01/2013 12:00		11/01/2013 12:00		
P9-1				23/08/2012 12:00	24/08/2012 5:00		
P10-1			29/08/2012 11:00			01/09/2012 11:00	
P10-2				07/03/2013 12:00		09/03/2013 12:00	
P11-1	08/10/2012 12:00				05/10/2012 12:00		

Tabla 25: Resumen de la distribución temporal de las mediciones de presión disponibles. Las casillas coloreadas de azul corresponden a la época otoño-inverno y de naranja a la época primavera-verano.

- Se calcula la media de las 6 mediciones realizadas por hora (recordar que la toma de datos es cada 10 min).
- El promedio de las mediciones que no superan en 0.6 mca la media calculada en el anterior paso, es el valor que se considera representativo de cada hora.
- En aquellos casos en los que había varios registros coincidentes en la misma época del año y día de la semana, se ha tenido en cuenta la media de los mismos.

Los gráficos incluidos en el Anejo 2 representan el resultado final obtenido, el cual se ha guardado en una hoja de cálculo a parte (*R5_Mediciones_Presion_Result_Def.xlsx*).

3.5. Modelación del sistema de abastecimiento

Tras pos-procesar la información disponible, queda evidenciado que podrá modelarse únicamente parte del sistema, en tanto que sólo existen datos de calibración asociados al sector 3. Excepto la simplificación de la red y asignación de cargas, el resto de procedimientos se efectúan exclusivamente para este sector. Si bien, una vez se consigan una cantidad mínima de datos para la calibración, se pueden aplicar esos mismos procedimientos al resto.

3.5.1. Simplificación de la red y empleo de GHydraulics para crear el INP de EPANET

La simplificación de la red queda ceñida únicamente a la supresión de los ramales, que se sustituyen por su punto inicial. El resto de tuberías, aun siendo de pequeño diámetro, se incluye en el modelo, ya que a priori no se puede evaluar su capacidad de transporte y, seguramente, en muchos casos no será despreciable, sobre todo porque pertenecen al entramado de mallas. El único ramal al que se le ha dado un tratamiento particular es el identificado en la **Figura 25**, que representa un área de notable magnitud, por lo que parecía acertada su inclusión en el modelo.

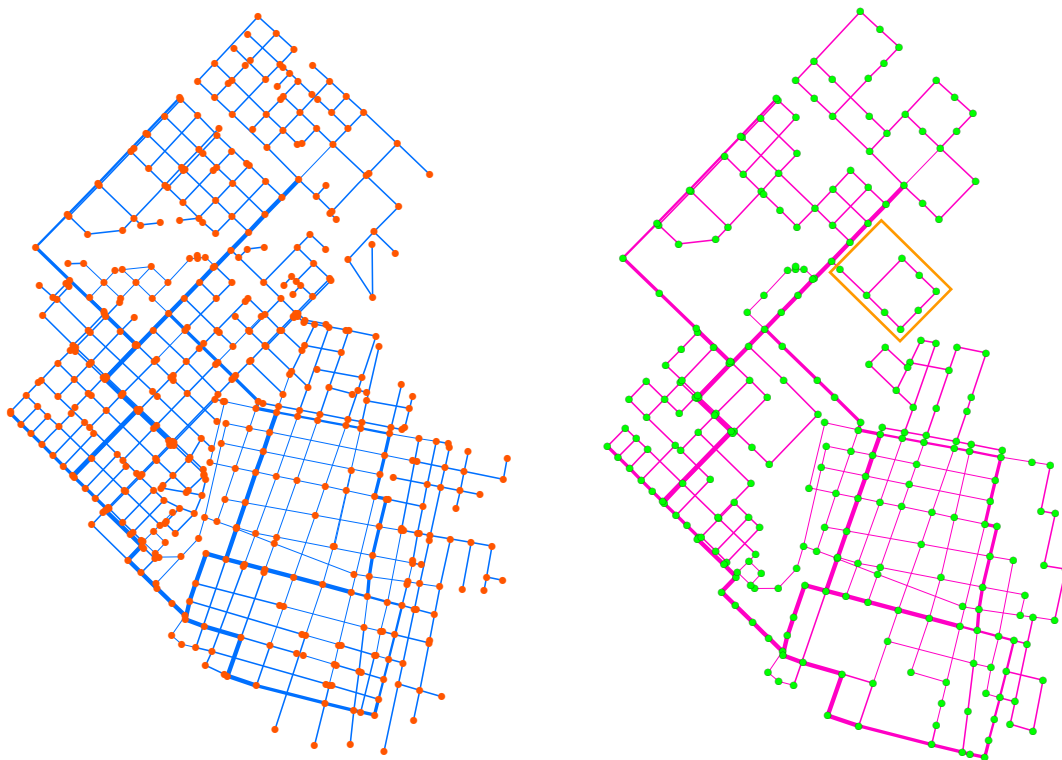


Figura 25: A la izquierda está el esquema completo de la red y a la derecha el simplificado, con el ramal incluido recuadrado.

La edición de las capas que contienen la solución final, *pipes* y *junctions*, se ha llevado a cabo mediante ArcGIS y parten de *512_510_GHydr_TubORI* y *513_511_GHydr_NudORI*, respectivamente. Los campos que poseen las mismas se han modificado de tal manera que respetan el formato demandado por GHydraulics, extensión de QGIS empleada para transformar las capas SIG de puntos y tuberías a un archivo INP de EPANET. El resultado obtenido al aplicar dicha extensión se guarda con el nombre *M1.inp*.

Las capas SIG donde se encuentra la simplificación propuesta para los sectores 1 y 2 son *515_513_Simpl_Nud* y *514_512_Simpl_Tub*.

3.5.2. Puntos de producción

El sector 3 está alimentado a través de dos puntos de producción, tal y como se ha expuesto en el apartado 3.2.

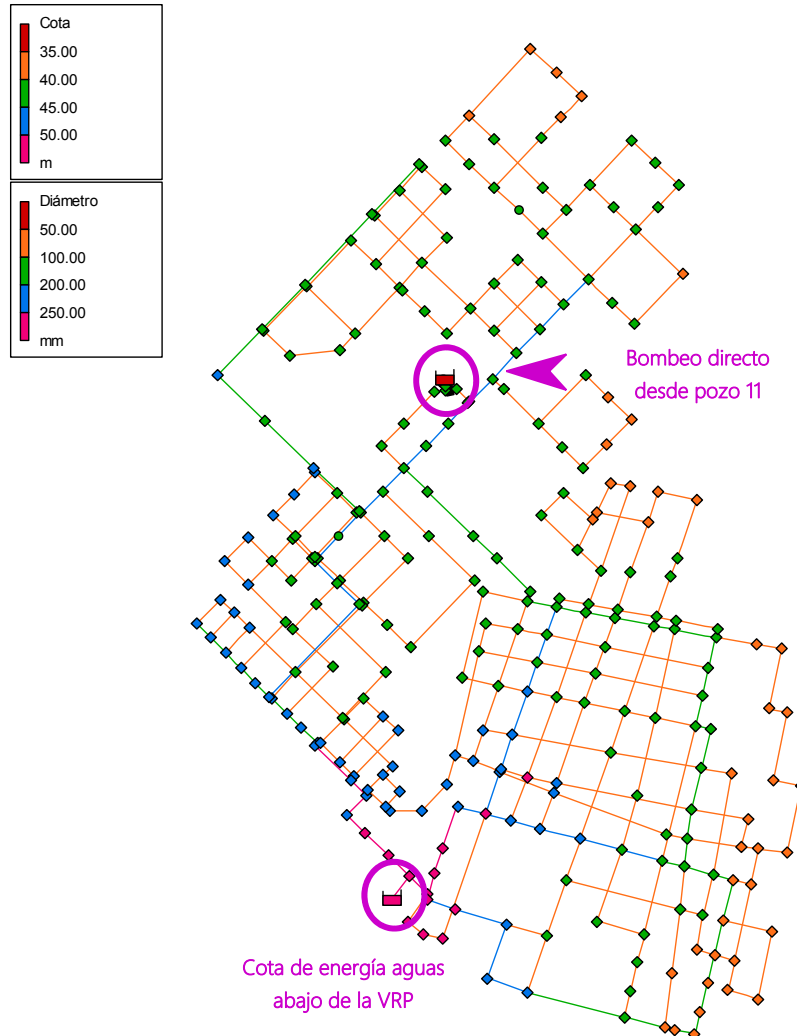


Figura 26: Esquema del modelo introducido en EPANET con los puntos de producción señalados.

El punto de producción constituido por el tanque principal nro. 35, junto a los 7 pozos que bombean agua hasta el mismo, es el de mayor importancia. De éste son desconocidos los datos acerca de las tuberías de conexión entre pozos y tanque y entre tanque y red, así como la configuración en época estival asociada a la automatización que controla el arranque y parada de las bombas en función del nivel de dicho tanque. El caudal inyectado medido corresponde precisamente a esta época. En tanto que se dispone de la modulación de la cota de energía aguas abajo de la VRP instalada a pie de tanque, se decide modelar todo el conjunto como un punto de inyección compuesto por un embalse a cota 65.72 m, al que se le aplica un patrón de alturas ($Id = E7$) que reproduce la modulación mostrada en la Figura 24. Éste se conecta directamente a la red mediante una tubería ficticia ($Id = TE$), que

no genera ningún impacto en el comportamiento del modelo, ya que su diámetro es de 1000 mm y su longitud de 1 m.

El otro punto de producción está formado por el tanque 34 y el pozo 11. Tal y como se explicó en el apartado 3.2, normalmente se inyecta agua a red desde el pozo 11 a través de una conexión directa. Es por ello que sólo se ha incluido en el modelo la misma. Con respecto a la bomba, es conocido que funciona 24 horas al día y la profundidad a la que está instalada, 26 m, pero no la curva de descenso dinámico del nivel de agua en el pozo, siquiera el nivel estático. Por otro lado, la referencia proporcionada de la misma no concuerda con ninguna de las incluidas en el catálogo del proveedor, *Rotor Pump*. Tanto dicha referencia incompleta (**E6**), como la potencia contratada (15 Kw) y el punto de funcionamiento dato ($50 \text{ m}^3/\text{h}$, 55 mca), sugieren que se trata del modelo RP6S55-6Q, pese a que ello contradiga la potencia dato facilitada (5.5 kw). Finalmente, se decide introducir en el modelo un embalse a 26 m por debajo del nivel del terreno (el cual está a 44.5 msnm), y enlazarlo mediante una bomba a la tubería de conexión directa. La curva característica de la misma corresponde a la del modelo RP6S55-6Q de *Rotor Pump* (<http://www.rotorump.com.ar/pdf/RP6S.pdf> 8933.pdf).

3.5.3. Asignación de cargas

3.5.3.1. Consumo contabilizado

El objetivo perseguido en este apartado es determinar los nudos de consumo del modelo, así como establecer para cada uno de ellos, a partir de los datos procedentes de María Grande, una función de densidad de la demanda base asociada, la cual permite definir la probabilidad de que dicha demanda se encuentre dentro de cierto rango.

3.5.3.1.1. BLOQUE 5 del proyecto SIG: Demanda

En primer lugar, se abre el BLOQUE 5 en el proyecto SIG iniciado, específico para el desarrollo de este apartado, al que se le ha procurado el nombre de *Demanda*. Su propósito es designar un número de nudos de consumo tal que a cada uno de ellos le corresponda como mínimo 20 parcelas. Se escoge este umbral en un intento de alcanzar una solución de compromiso entre:

- Una distribución uniforme de nudos de consumo a lo largo de la red y
- Un nivel de agregación de consumos que permita suponer una modulación suave, no afectada por la aleatoriedad intrínseca a la demanda.

El tipo de asignación por la que se ha optado sería una combinación entre las realizadas por mallas y zonas. En concreto, se han distinguido áreas de influencia que comprenden una malla, sus nudos y los ramales que parten de la misma. A su vez, dichas áreas cubren un número determinado de manzanas, en las que se ubican parcelas de distinta clase. Es decir,

las áreas de influencia delimitadas constituyen la entidad espacial empleada para relacionar parcelas con nudos de consumo.

El resultado obtenido lo constituyen los archivos descritos en la **Tabla 26**. Los pasos seguidos para alcanzarlos se explican a continuación.

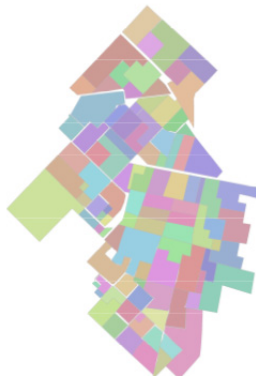
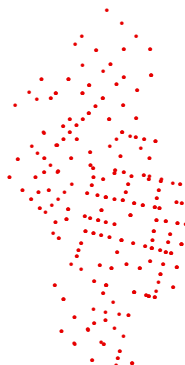

Nombre de la capa	Descripción	
902_AreasInfluenciaMallas1	Capa de polígonos que representan las áreas de influencia que permiten asignar parcelas a nudos de consumo.	
907_515_Edit_NudosConsumo	Capa de puntos que incluyen los nudos de consumo propuestos para el modelo de la red.	
516_515_910DBF_SimplNudDemand	Capa que incluye todos los puntos de la red simplificada, en la que los nudos de consumo disponen de su nuevo identificador y se puede consultar el número de parcelas de cada clase que les corresponden.	

Tabla 26: Resultados obtenidos en el Bloque 5 del SIG, Demanda.

Paso 1: Cálculo de la cantidad de parcelas de cada clase ubicadas en cada manzana.

Esta operación se lleva a cabo en una hoja de cálculo y parte del resultado alcanzado en el apartado 3.4.1.2.2. Para cada manzana se calcula el número de parcelas de cada clase mediante una suma condicionada. La tabla resultado se une a la capa de polígonos *105_104_708_SpatialJoinEditTableText_Manz* a través del campo que tienen en común,

ID_Uniform. Dicha unión se hace permanente guardando la capa con el nombre *901_105_EditTableIDUni_DemandManz*.

Paso 2: Edición de las áreas de influencia

Se crea una capa de polígonos, llamada *902_AreasInfluenciaMallas1*, a partir de las líneas incluidas en el esquema simplificado de la red (*515_513_Simpl_Nud*). Se termina de completar la misma modificando o añadiendo polígonos para que todas las manzanas con parcelas clasificadas queden incluidas dentro de alguna área. Su delimitación tiene en consideración que:

- La margen de la parcela que es colindante con una calle indica que la vivienda ubicada en ella está conectada a la tubería que discurre por dicha calle.
- Los ramales deben incluirse dentro del área que comprende la malla de la que parten. Cuando no es posible, se intenta compensar unas zonas con otras, de forma que, en orden de magnitud, a todas ellas les corresponda un número de parcelas acorde a aquellas que abastecen.

Paso 3: Selección de los nudos de consumo

A partir de los nudos de la capa *515_513_Simpl_Nud*, se crea otra, *907_515_Edit_NudosConsumo*, que sólo contiene aquellos que se consideran de consumo. Estos son seleccionados buscando una distribución espacial uniforme a lo largo de la red. También se intenta respetar los nudos de inicio de los ramales. A su vez, las áreas de influencia se editan conjuntamente, con el objetivo de que sólo estén en contacto con los nudos de consumo correspondientes. Por otro lado, se añade un nuevo campo que se cumplimenta con el nuevo identificador que les corresponde a los nudos por ser de consumo, acabado en *01* en vez de en *02* (nudos de conexión). Con la asistencia de una hoja de cálculo (*R6_ID_Nud.xlsx*), se han actualizado dichos identificadores en las capas *515_513_Simpl_Nud* y *514_512_Simpl_Tub*, añadiendo también nuevo campos.

A priori no es posible saber si el resultado obtenido en este paso cumple con el umbral establecido de 20 parcelas por nudo de consumo. Así pues, tanto ésta como las siguientes operaciones forman parte de un proceso iterativo que acaba en el momento quede verificada esta condición. Finalmente, el número de nudos de consumo pertenecientes al sector 3 asciende a 160.

Es de importancia añadir que existen 2 nudos, el *305201* y *328001*, a través de los cuales se pretende tener en cuenta el abastecimiento de aquellas parcelas fuera de la planta urbana consolidada, de las que no se dispone de información, pero están conectadas a la red pública de agua. En próximos apartados se hará referencia a su tratamiento.

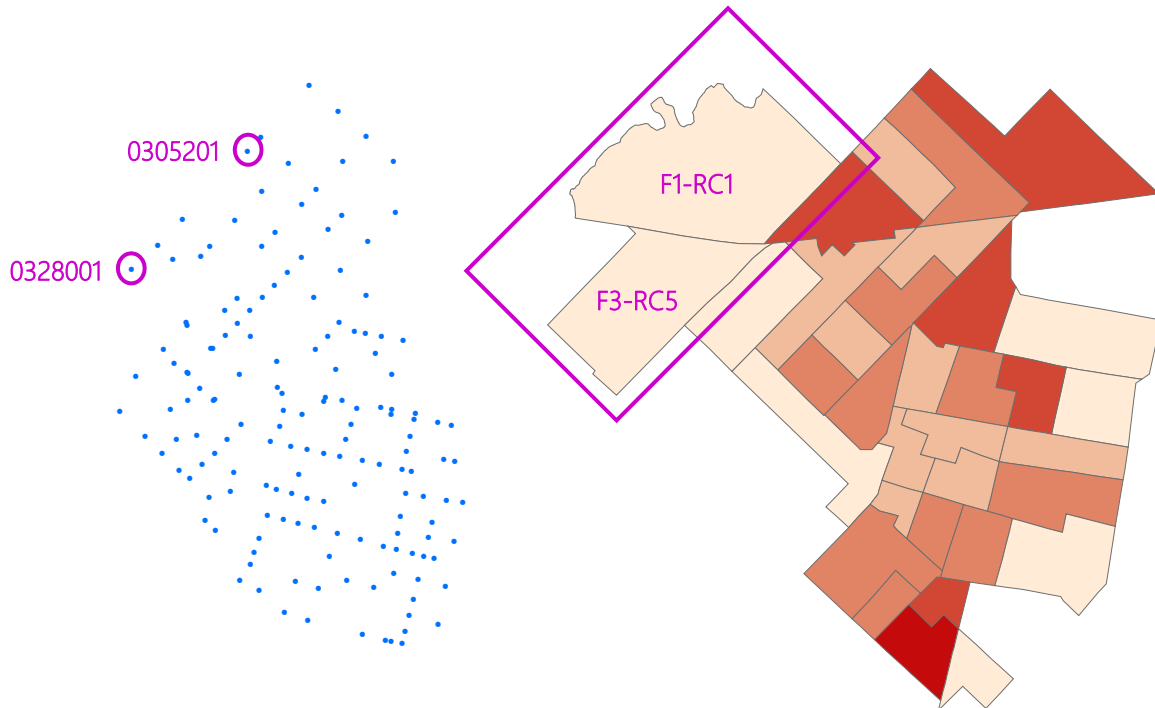


Figura 27: Identificación de los nudos 305201 y 328001 a través de los cuales se quiere tener en cuenta el consumo de las parcelas ubicadas fuera de la planta urbana consolidada, que quedarían comprendidas en los radios censales 1 y 5, pertenecientes a las fracciones 1 y 3, respectivamente.

Paso 4: Cálculo del número de parcelas de cada tipo y el número de nudos de consumo por área de influencia

Se unen los datos procedentes de la capa *901_105_EditTableIDUni_DemandManz* a la capa *902_AreasInfluenciaMallas1* basándose en la localización espacial. Al seleccionar la opción de resumir dichos datos mediante la operación suma, se consigue saber el número de parcelas de cada clase que incluye un área de influencia. Con ello se obtiene la capa de polígonos *908_901_902_Join_ArealInfluNudoConsNroViv*.

Al resultado anterior se aplica de nuevo la misma herramienta, pero en esta ocasión los datos proceden de *907_515_Edit_NudosConsumo*. No se selecciona ninguna opción de resumen, en tanto que se busca únicamente contabilizar los nudos de consumo en contacto con las áreas de influencia. En la capa obtenida, guardada como *909_907_908_Join_NudoConsPorArealInflu*, se crean tantos campos como clases de parcela, donde se calcula la cantidad de las mismas que corresponde a los nudos de consumo en contacto con un área de influencia concreta. Esto se determina dividiendo la información originada en *908_901_902_Join_ArealInfluNudoConsNroViv* según el campo creado en la operación anterior.

Paso 5: Cálculo de las parcelas de cada tipo por nudo de consumo

Por último, se emplea la herramienta *Analysis Tools > Overlay > Spatial Join* a las capas *907_515_Edit_NudosConsumo* y *909_907_908_Join_NudoConsPorAreaInflu*, consiguiéndose como resultado la capa *910_907_909_SpatialJoin_RelIDNudAreaInflu*. Ésta contiene los nudos de consumo repetidos tantas veces como áreas de influencia estén en contacto con los mismos. A través de la herramienta *Summarize* se obtiene una tabla DBF en la que para cada nudo de consumo están calculadas el número de parcelas de cada clase que le corresponde. La tabla generada se une por identificadores comunes (el ID del nudo) a la capa *907_515_Edit_NudosConsumo*. Para hacer esta unión permanente se guarda la capa con el nombre *516_515_910DBF_SimplNudDemand*.

3.5.3.1.1. Clústering de nudos de consumo

Los nudos pertenecientes al sector 3 se almacenan por separado en *c2_nudcons_01_M1_516_515_910DBF_SimplNudDemand*. Los valores de los campos donde están registradas las parcelas de cada clase, coligadas a los nudo de consumo, se redondean, con excepción de los correspondientes a la clase 6, la cual contiene las parcelas con un consumo de agua asociado muy superior al del resto de clases. Éstos se editan con especial cuidado: las parcelas de dicha clase que aparecen en cada manzana se asocian únicamente a los nudos de consumo más cercanos. Esta operación se realiza manualmente.

Seguidamente, se prepara una tabla de datos con 160 filas, una por cada nudo de consumo, y 6 columnas, una por cada clase de parcela, que sirve como entrada a un clústering de nudos de consumo. El propósito del mismo es encontrar el tipo de clústering, métrica y la cantidad de agrupaciones que minimicen las diferencias en cuanto a número de parcelas de cada clase entre individuos del mismo clúster. El coeficiente de variación calculado por cada grupo y clase es el indicador escogido para evaluar las distintas soluciones obtenidas (*R6_analisis_cv.xlsm*).

Los tipos de clústering empleados coinciden con los del clústering de parcelas expuesto en el apartado 3.4.1.2.1.5. En cambio, no se han testeado las métricas Euclidiana y Manhattan aplicadas a las variables de entrada tipificadas, dado que en este caso dichas variables son del mismo orden de magnitud. La cantidad de agrupaciones tenidas en cuenta son 10, 20 y 30. De este modo el total de pruebas realizadas suman 51.

Los gráficos que aparecen en las **Figuras 28, 29 y 30** muestran el coeficiente de variación por prueba realizada y clase de parcela para 10, 20 y 30 clústeres, respectivamente. Para cada cantidad de agrupaciones, los tipos de clústering que minimizan el coeficiente de variación, ordenados de mejor resultado a peor, son:

Número de agrupaciones 10

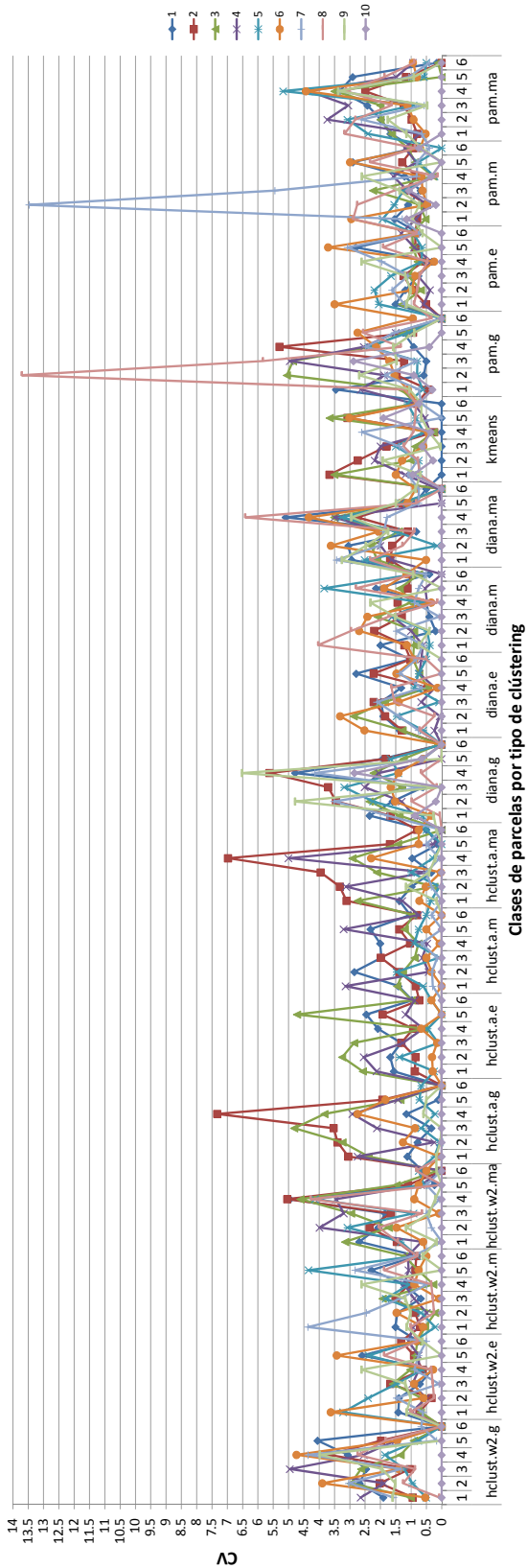


Figura 28: Representación del CV por clase de parcela y prueba de clústering realizada, para un 10 clústeres.

Número de agrupaciones 20

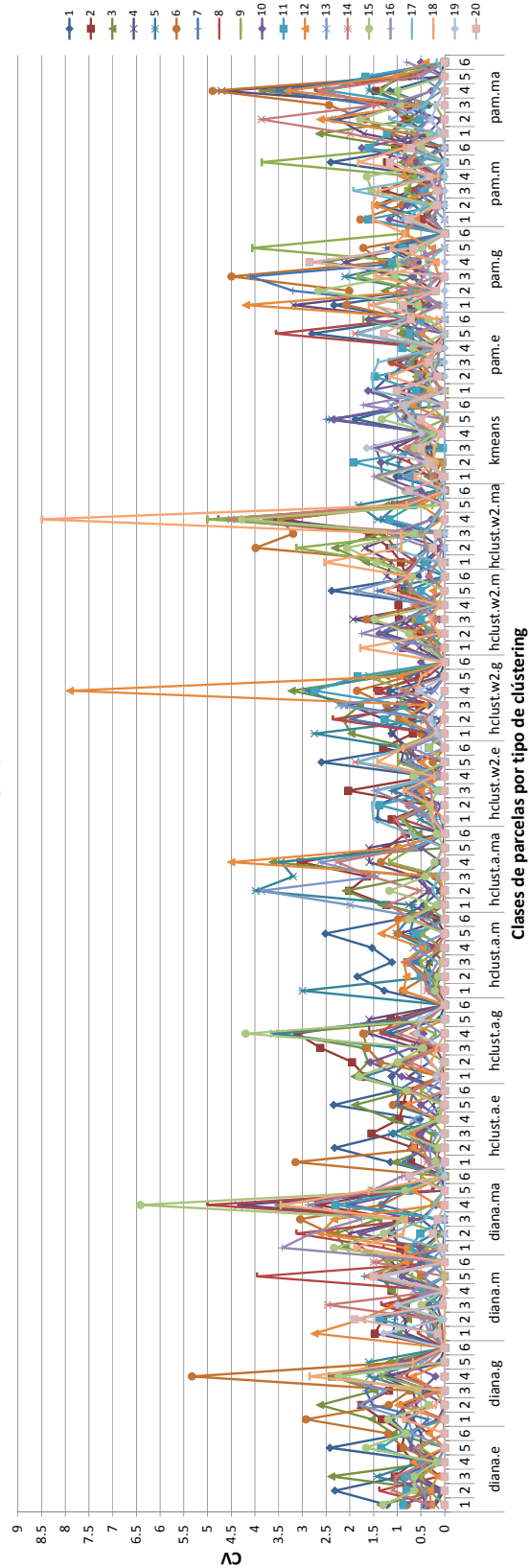


Figura 29: Representación del CV por clase de parcela y prueba de clústering realizada, para un 20 clústeres.

Número de agrupaciones 30

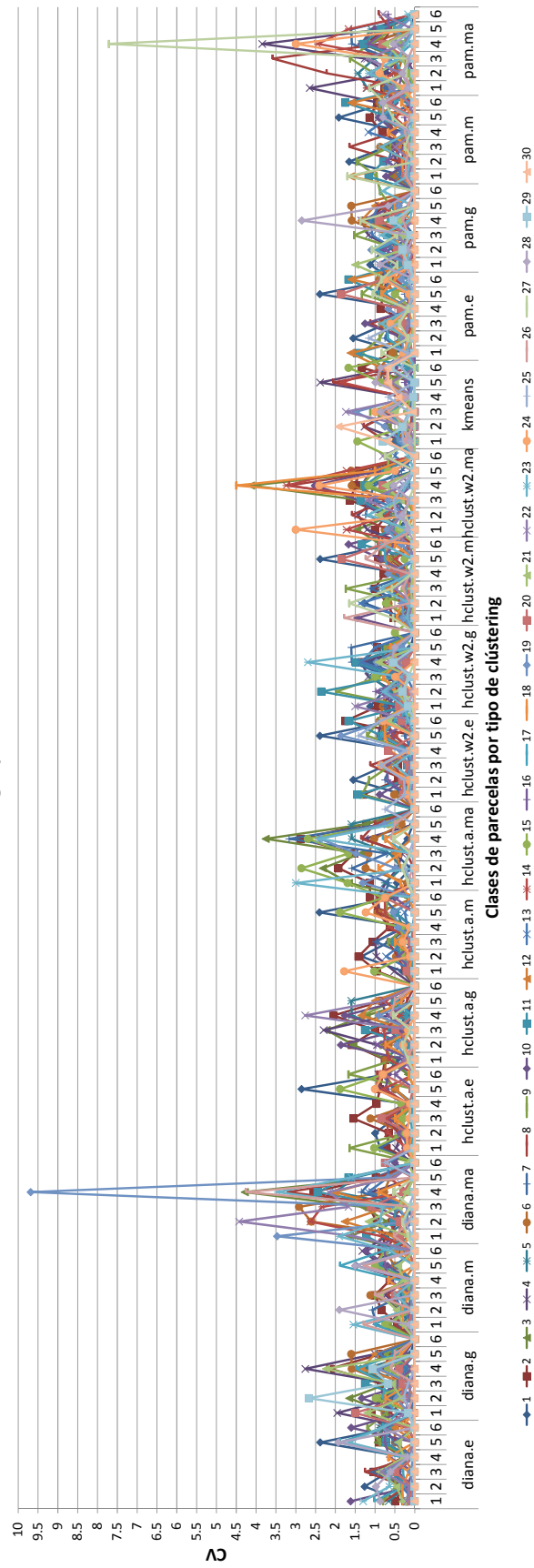


Figura 30: Representación del CV por clase de parcela y prueba de clústering realizada, para un 30 clústeres.

- 10 clústeres:
 - *diana.e*: clústering jerárquico divisivo, métrica Euclidiana.
 - *hclust.a.m*: clústering jerárquico aglomerativo, método UPGMA, métrica Manhattan.
- 20 clústeres:
 - *hclust.w2.e*: clústering jerárquico aglomerativo, método Ward.D2, métrica Euclidiana.
 - *hclust.w2.m*: clústering jerárquico aglomerativo, método Ward.D2, métrica Manhattan.
 - *hclust.a.m*: clústering jerárquico aglomerativo, método UPGMA, métrica Manhattan.
 - *diana.e*: clústering jerárquico divisivo, métrica Euclidiana.
 - *Kmeans*: clústering de partición KMEANS.
- 30 clústeres:
 - *pam.m*: clústering de partición PAM, métrica Manhattan.
 - *diana.m*: clústering jerárquico divisivo, métrica Manhattan.

Finalmente, la solución escogida es la correspondiente a *30 agrupaciones* y al *clústering de partición PAM, métrica Manhattan*, ya que el coeficiente de variación se reduce considerablemente y sólo hay 7 clústeres con un número de individuos igual a 3 o menor.

Por último, en tanto que se desconoce el número de parcelas fuera de la planta urbana consolidada que se abastecen de la red pública de agua, así como sus características, se decide crear 2 grupos más, una para *305201* y otro para *328001*. El resultado definitivo se puede ver en la **Figura 31**.

LEYENDA

cl_pamm30

- 0
- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9
- 10
- 11
- 12
- 13
- 14
- 15
- 16
- 17
- 18
- 19
- 20
- 21
- 22
- 23
- 24
- 25
- 26
- 27
- 28
- 29
- 30
- 31
- 32

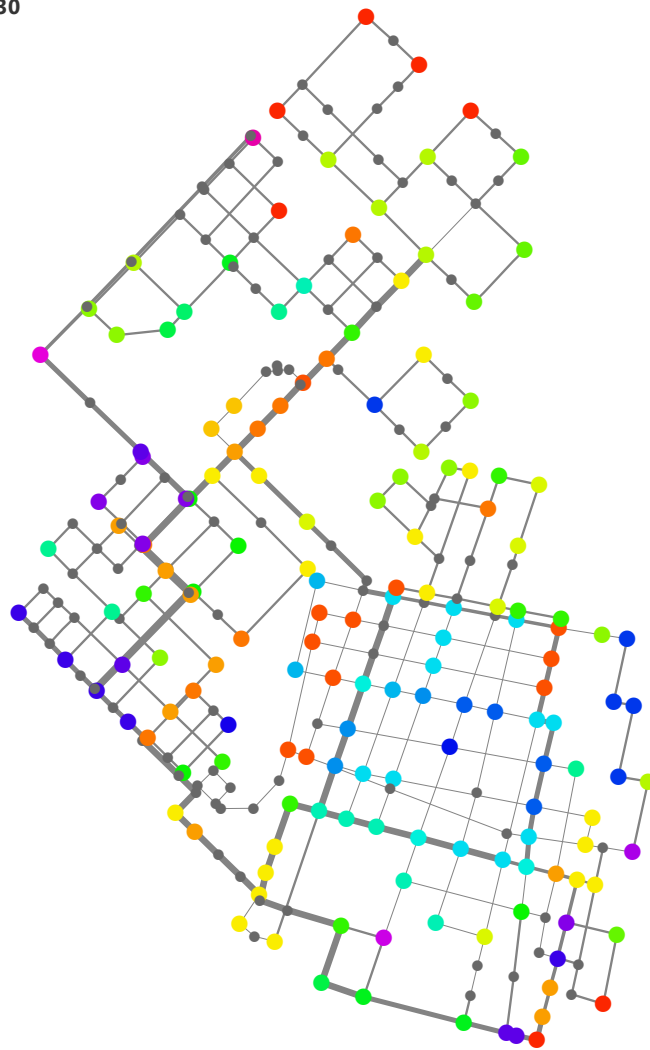


Figura 31: Resultado final del clústering de nudos.

3.5.3.1.2. Generación aleatoria de una muestra de demandas probables para cada grupo de nudos de consumo y cálculo de su correspondiente función de densidad

La idea principal detrás del siguiente procedimiento es generar una muestra aleatoria de observaciones simuladas de demanda base de cada variable *nudos de consumo de la clase i*, con $i=1, \dots, 30$ (evidentemente, no se puede llevar a cabo en el caso de los nudos 305201 y 328001), tomando como base los datos de María Grande. Posteriormente, se estima la probabilidad de que dicha demanda base en los distintos grupos se encuentre dentro de un cierto rango.

Así pues, se prepara una tabla donde se calcula un valor medio redondeado representativo de la cantidad de parcelas por clase para cada grupo de nudos de consumo

(*R6_analisis_cv.xlsm*). La misma es el dato de entrada de una subrutina escrita en una hoja de cálculo, que lleva a cabo los siguientes pasos para cada grupo:

1. Se genera tantos números aleatorios como cantidad de parcelas por clase, asegurando que dentro de cada una de ellas no se repiten dichos número aleatorios. El rango posible de estos últimos está acotado por la cuantía de parcelas disponibles de cada clase pertenecientes a María Grande.
2. Para ese conjunto de números aleatorios se obtiene la suma de los consumos asociado a las parcelas seleccionadas de María Grande.
3. Se registra el resultado obtenido en el paso anterior.

Con dicha subrutina se ha generado una muestra de 1000 realizaciones para cada variable *nudos de consumo de la clase i*. No se puede asegurar que en el conjunto de las mismas se hayan elegido todas las parcelas de María Grande con la misma frecuencia.

El anterior resultado se emplea en la estimación no paramétrica de la función de densidad de cada variable *nudos de consumo de la clase i*. El tipo de estimación empleada es la conocida como Núcleo o Kernel, implementada en la función *density* de R Statistics. Dicha estimación depende de dos parámetros: la función Kernel y el ancho de ventana (también llamado parámetro de suavizado o *bandwith* en inglés). Se ha escogido la Función Kernel Gaussiana, ya que es la empleada con mayor frecuencia. Por otra parte, se ha optado por la configuración de la función *density* que usa los métodos de Sheather & Jones (1991) asociados éstos a la Función Kernel Gaussiana (otra razón que justifica su elección), para encontrar un ancho de ventana óptimo a partir de los datos. El resultado obtenido se muestra en la **Figura 32**.

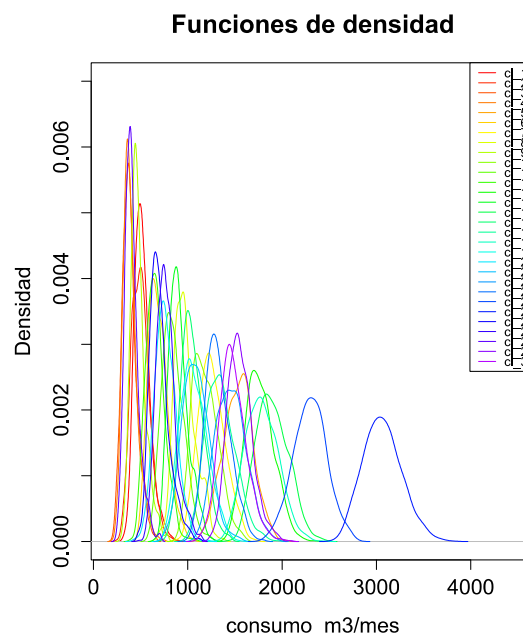


Figura 32: Funciones de densidad correspondientes al consumo por clase definida en el clústering de nudos de consumo.

3.5.3.2. Consumo no contabilizado

Para este trabajo el consumo no contabilizado se reduce a las fugas en la red. Éstas se han modelado mediante los emisores de EPANET, es decir, se han considerado dependientes de la presión. El exponente del emisor se ha dejado con el valor por defecto, 0.5, mientras que los coeficientes de descarga se ajustaran durante la calibración del modelo. En el apartado 3.5.5 se concluye el tratamiento que se les procura a los coeficientes de descarga.

3.5.4. Asignación de mediciones de presión a nudos del esquema simplificado

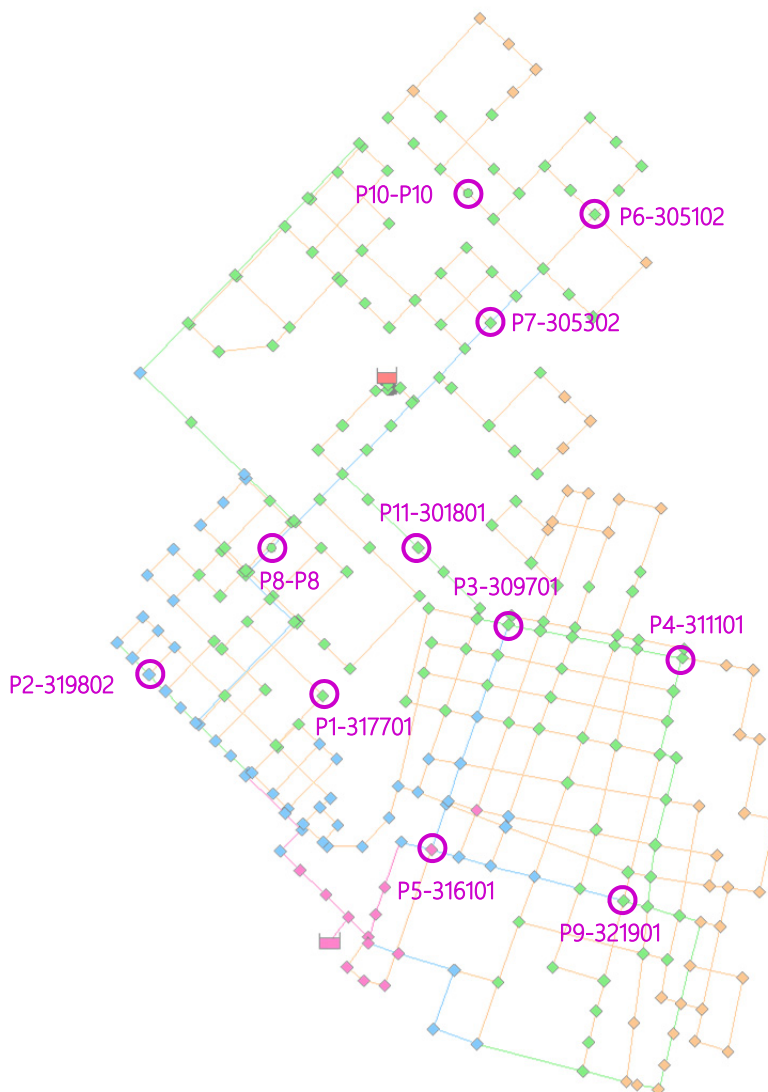


Figura 33: Ubicación final en el modelo de los registros disponibles de presión medida.

Las ubicaciones donde se colocaron los medidores de presión se han proporcionado con referencias no precisas. Es por ello, que en general se asocian al nudo más cercano del

esquema simplificado de la red, intentado asegurar de forma aproximada que la distancia máxima entre la posición real y la adjudicada no supera los 25 m. Para cumplir esto último, ha sido necesario crear en las tuberías 322903 y 336003 los nudos ficticios, P10 y P8, respectivamente (**Figura 33**), los cuales sólo han sido introducidos en el modelo de EPANET M1. Por desdichado, se ha respetado el dato acerca del diámetro de la tubería distribuidora, a la que directamente estaba conectada la acometida sobre la que se instaló el medidor de presión.

Los registros P7 y P9 están asociados a mediciones comprendidas en la época de invierno-otoño, así pues no se utilizarán como datos de calibración. Por otro lado, el registro P1 se ubica en un ramal y está notablemente alejado del punto inicial del mismo, razón por la cual tampoco se ha empleado. No obstante, la información que proporcionan todos ellos permitirá observar indicios que sirvan para calificar aproximadamente la bondad del modelo, por lo que se les adjudicará el punto más cercano de éste.

Es de importancia aclarar que todos los nudos (excepto P8 y p10 que se han añadido para enlazar al modelo los datos de calibración), es decir, tanto aquellos que son de conexión como de consumo en el esquema simplificado, son considerados potenciales emisores.

3.5.5. Clústering de tuberías y nudos

En un intento de reducir el número de grados de libertad en el proceso de calibración, se ha realizado un clústering de tuberías relacionado con su rugosidad y uno de nudos tocante a los coeficientes de descarga. Ambos se basan en la idea de que elementos de características parecidas sometidos a condiciones similares se comportaran de forma análoga. Las variables empleadas son:

3. Clústering de tuberías:
 - *Coordenadas del punto medio de la tubería* (2 variables numéricas)
 - *Material, edad y diámetro* (2 variables nominales y 1 numérica, respectivamente)
4. Clústering de nudos:
 - *Coordenadas y cota* (3 variables numéricas)
 - *Longitud de tubería ponderada*, considerando el esquema completo, y *densidad de acometidas* (2 variable)

La elaboración de los datos necesarios en el caso del clústering de tuberías es prácticamente directa a partir de la capa 514_512_Simpl_Tub (apartado 3.5.1), de la cual se aprovecha la información referente al sector 3. Las coordenadas del punto medio se calculan a través de la *Calculadora de Campo* aplicada a la capa c2_tub_01_vcoord_M1pipes, obtenida mediante la herramienta *Data Management Tools > Features > Feature To Point*. En el caso del clústering de nudos, mientras que las variables *coordenadas* y *cotas* ya están implícitas o disponibles en *junctions* (apartado 3.5.1), la *longitud de tubería ponderada* y la

densidad de acometidas requirieron mayor carga de trabajo. Los pasos seguidos para conseguirlos son:

Longitud de tubería ponderada

Los datos de partida corresponden a la información sobre el sector 3 contenida en la capa *512_510_GHydr_TubORI*, que almacena el esquema completo de la red, los cuales se guardan en una nueva *c3_cdesc_01_vlong_512_TubOri*. Se crean 3 nuevos campos en esta última con ArcGIS, que se cumplimentan gracias a la selección espacial:

Id_ramal: A cada ramal se le adjudica un identificador y el mismo es asociado a las tuberías que pertenecen a un determinado ramal.

Id_tubc1: Haciendo uso del mismo identificador, se señalan las tuberías que son cabecera de ramal.

Id_tubrs: Este campo toma el valor 0 cuando la tubería no pertenece al modelo simplificado y 1 en caso contrario.

Tomando como entrada la tabla editada en *c3_cdesc_01_vlong_512_TubOri* y la correlación entre identificadores de nudo del esquema completo y el esquema simplificado, es determinada con la asistencia de una hoja de cálculo (*R7_c3_dat_resultclus.xlsx*) la longitud ponderada que corresponde a cada nudo.

Densidad de acometidas

Los pasos seguidos son análogos a los expuestos en el apartado 3.5.3.1.1, en el que se describe el trabajo realizado en SIG para asignar parcelas a nudos de consumo, con la diferencia de que en este caso sólo se ejecutan las operación para el sector 3, interviene todos los nudos del esquema simplificado (excepto P8 y P10) y el número de acometidas parte del número de parcelas con superficie construida mayor que 0. El resultado final se guarda en la tabla DBF *c3_cdesc_04_vnroaxnud*.

Por otro lado, en ambos clústering existen datos con diferente grado de fiabilidad, incertidumbre o importancia. Esta es la razón por la que sólo se ha empleado la métrica Gower, la cual permite adjudicar distintos pesos a cada variable de entrada que interviene y no se ve afectada por el orden de magnitud de las mismas. Además, en el caso del clústering de tuberías, la edad y el material son variables nominales, por tanto, únicamente puede utilizarse dicha métrica.

Después de llevarse a cabo algunos tanteos, se opta por considerar los tipos de clústering, combinación de pesos y cantidad de agrupaciones incluidas en la **Tabla 27**. Se prueban todas las posibles combinaciones entre ellos.

Clústering de tuberías		
Tipos de clústering	Combinación de pesos	Nro. de clústeres
Clústering jerárquico aglomerativo. Método ward.D2.	Sin pesos	15
Clústering de partición PAM	Comb_w1: $0.3 * \text{Diámetro} + 0.3 * \text{Edad} + 0.3 * \text{Material} + 0.05 * \text{Coordenada X} + 0.05 * \text{Coordenada Y}$	20
	Comb_w2: $0.3 * \text{Diámetro} + 0.25 * \text{Edad} + 0.25 * \text{Material} + 0.1 * \text{Coordenada X} + 0.1 * \text{Coordenada Y}$	25
Clústering de nudos		
Tipos de clústering	Combinación de pesos	Nro. de clústeres
Clústering jerárquico aglomerativo. Método ward.D2.	Comb_w1: $0.3 * \text{Cota} + 0.3 * \text{longitud ponderada} + 0.15 * \text{Coordenada X} + 0.15 * \text{Coordenada Y} + 0.1 * \text{Nro. acometidas}$	10
Clústering jerárquico aglomerativo. Método UPGMA.	Comb_w2: $0.3 * \text{Diámetro} + 0.2 * \text{Edad} + 0.15 * \text{Material} + 0.15 * \text{Coordenada X} + 0.2 * \text{Coordenada Y}$	20
Clústering jerárquico divisivo DIANA		30
Clústering de partición PAM		

Tabla 27: Tipos de clústering, combinación de pesos y número de clústeres combinados en las pruebas llevadas a cabo de clústering de tuberías y parcelas.

En ambos clústering se comprueba que el número máximo de agrupaciones testado parece el más adecuado. En el caso de las tuberías porque se consigue generar un grupo en cada tamaño de diámetro, así como entre materiales y edades presentes. Para los pequeños diámetros se distingue una cantidad suficiente de clústeres según la localización espacial y material. En el caso de los nudos, se alcanza un tamaño medio de grupo balanceado, es decir, no existe ningún clúster predominantemente mayor en cuantía de individuos en referencia al resto. Los resultados para distintos tipos de clústering, combinaciones de peso y dicho número máximo de agrupaciones, se pueden ver en el Anejo 3.

Finalmente, se selecciona el *clústering de partición PAM con combinación de pesos 1* para las tuberías, por ser el que mejor distingue entre diferentes diámetros, y el *clústering jerárquico aglomerativo, método ward.D2 y con combinación de pesos 2*, para los nudos, por ser el que más respeta la topología de la red. No obstante, en este último, obedeciendo siempre a las diferencias establecidas en la clasificación previa, se han realizado correcciones añadiendo nuevos grupos en todos aquellos casos donde se identifica una incongruencia grave asociada a que las agrupaciones no respetan precisamente la topología de la red, es decir, nudos que el algoritmo clasifica dentro del mismo clúster por estar cerca en el espacio, pero que no guardan ningún otro tipo de relación, pues pertenecen a zonas de la red distintas. En la

Figura 34 se muestran los resultados finales.

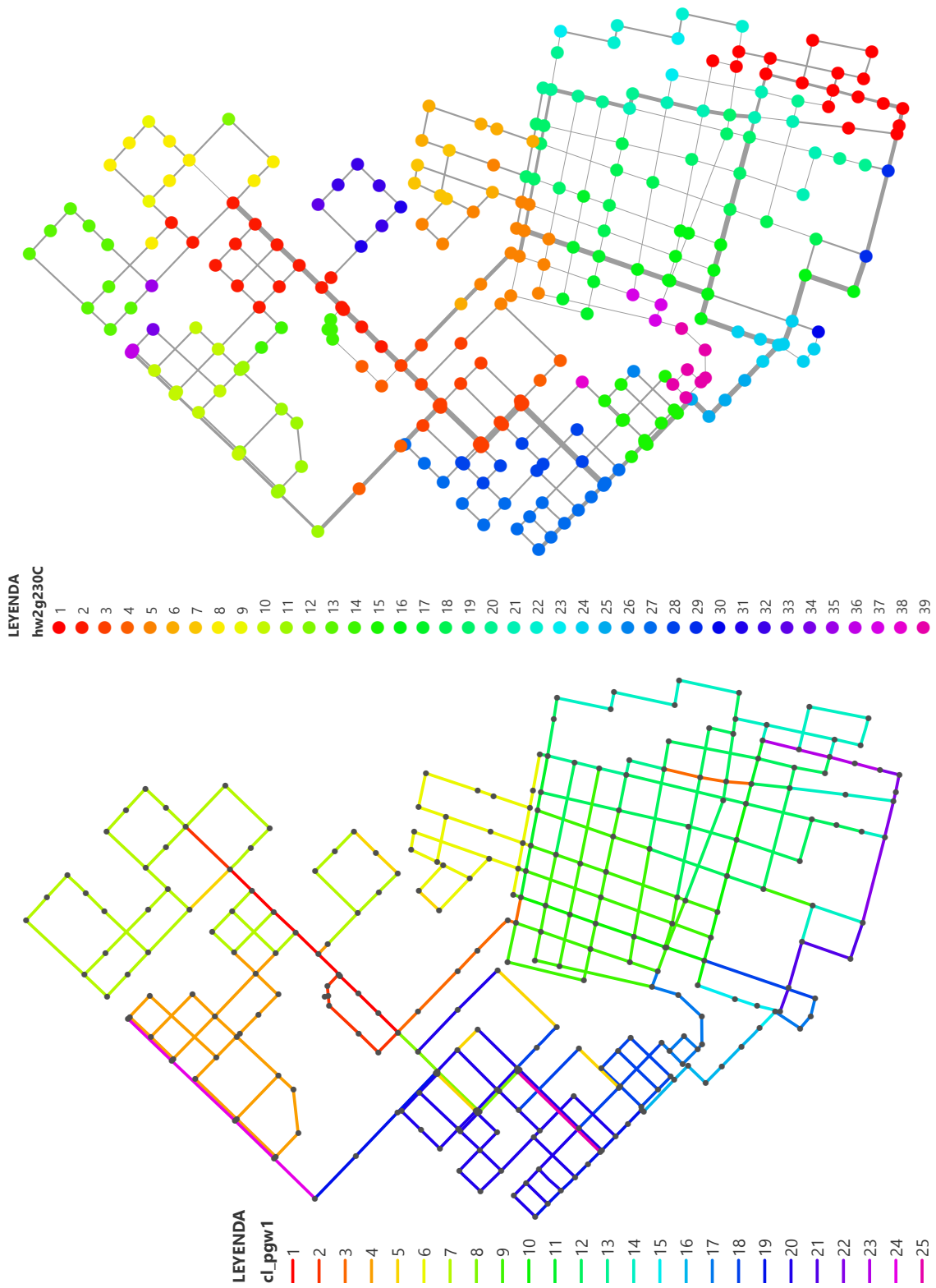


Figura 34: Resultado final del clústering de tuberías (abajo) y nudos (arriba).

3.6. Calibración

En los apartados anteriores se han empleado distintas técnicas para la elaboración del modelo del sistema de distribución de agua de Nogoyá. Si bien, cualquier modelo debe ser calibrado para que pueda afirmarse que es representativo de la realidad. Por consiguiente, se procede a realizar dicha operación, con el propósito de encontrar diferentes configuraciones del sistema que obtengan un valor de la función objetivo considerado como admisible.

El apartado 2.2 expone cómo está configurada la aplicación que permite calibrar el modelo combinando las herramientas Evolver y EPANET. La **Figura 35** detalla el proceso de calibración específico para el caso de estudio de Nogoyá. Así mismo, en los siguientes puntos se describen cada uno de los aspectos involucrados.

3.6.1. Datos de entrada del algoritmo

Evolver emplea un algoritmo evolutivo para la calibración del modelo. Los inputs que requiere son el propio modelo de la red, elaborado en EPANET, y la definición de la función objetivo, que parte de las mediciones de caudal y presión. Los términos de la misma constituyen funciones de penalización que hacen referencia a dichas mediciones, así como a la demanda base de los nudos (ver apartados 3.6.3.1 y 3.6.3.2).

3.6.1.1. Modelo de la red en EPANET

La organización del libro de cálculo (MS Excel) del que se nutre Evolver está basada en los datos del modelo de EPANET. En este sentido, el libro de Excel está conformado por distintas hojas de cálculo. Se dispone de dos específicas para los nudos de consumo y para las tuberías, en las que los elementos de la red se encuentran ordenados según el archivo INP del modelo. Esto es así para que en cada evaluación del mismo la configuración del sistema establecida por Evolver se introduzca correctamente en EPANET. Del mismo modo, existen otras dos hoja de cálculo centradas en las presiones observadas en el sistema y el caudal inyectado desde el tanque principal nro. 35, respectivamente. Se estructuran de forma que, una vez finalizada la simulación del modelo, se registren los resultados que corresponden a cada dato de calibración.

El libro de Excel también dispone de una hoja de *control*, en la que el algoritmo genético actualiza en cada iteración los valores de las variables de decisión (ver apartado 3.6.2). En este caso, las mismas se organizan según los resultados obtenidos en los distintos clústering realizados, lo cual facilita la introducción de sus límites en Evolver.

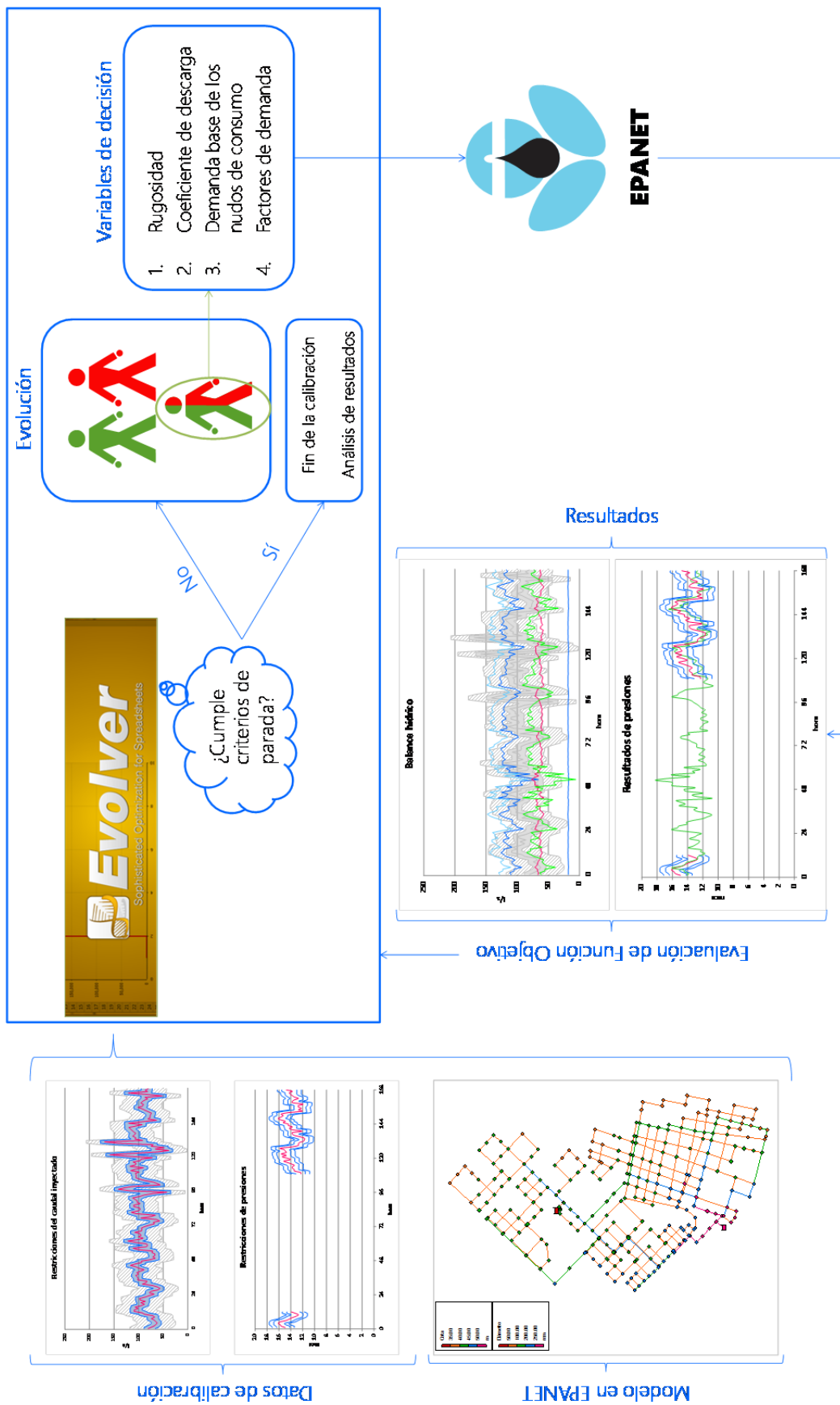


Figura 35: Esquema del proceso de calibración con Evolver de la red de suministro de agua de Nogoyá.

Para poder relacionar los valores de esta última hoja de *control* con los datos que se deben introducir en el modelo, se emplea el identificador de los distintos elementos. De esta manera (Figura 36), es sencillo buscar el valor adecuado en la hoja de *control* y registrarlo en la posición correcta en la que debe insertarse en Epanet.

Hoja de Control			Hoja de nodos de consumo	
Cl2_Nodos	ID_Nodos	Demanda base (l/s)	ID_Nodos	Demanda base (l/s)
1	300101	0.137478651	300101	0.137478651
1	305701	0.133657882	300301	0.342927134
1	305901	0.137151095	300401	0.151182035
1	306701	0.154634792	300501	0.102770628
1	307001	0.139175346	300601	0.238354727
1	322801	0.124316101	300701	0.115364219
1	324601	0.124210634	300801	0.450726697
2	300301	0.342927134	300901	0.728925524
2	309001	0.340620323	301501	0.208095728
2	309201	0.305938287	301602	0
2	309401	0.236741455	301702	0
2	311101	0.279154122	301801	0.637933617
2	311301	0.144968192	302101	0.527038629
2	311401	0.238608265	302202	0
2	312101	0.110676002	302301	0.126606396
2	345601	0.299351078	302402	0
2	345701	0.107369562	302501	0.40437288
2	345801	0.108124905	302901	0.438383098
3	300401	0.151182035	303002	0
3	300501	0.102770628	303101	0.511807243
3	300601	0.238354727	303301	0.484953704
3	308201	0.096896954	303402	0
3	310201	0.220368403	303801	0.1869613
3	318601	0.092152356	303901	0.218327945
3	320101	0.252631151	304201	0.468569967

Figura 36: Ejemplo de cómo buscar el valor adecuado en la hoja de control a partir del identificador del elemento.

3.6.1.2. Datos de calibración

La definición de dos términos de la función objetivo parte de los valores medios y umbrales admisibles del caudal inyectado y de las presiones observadas (Figura 37).

La obtención del valor medio observado ya se expuso en el apartado 3.4.2. Con origen en dichos valores se establecen dos umbrales: el primero define una banda tal que si los resultados quedan comprendidos en la misma son considerados como “aceptablemente correctos”; el segundo delimita otra banda más amplia, que se inicia a partir de la anterior, en la que los resultados se suponen “medianamente correctos”, mientras que más allá serían clasificados como “inadmisibles”. Estos dos umbrales determinarán las funciones de penalizaciones que componen los términos nombrados de la función objetivo (ver apartado 3.6.3).

Para establecer los umbrales y bandas se siguen los siguientes pasos:

- Se calcula la diferencia entre el valor observado y el medio correspondiente a cada hora y serie.

- Como primer umbral, se escoge el valor promedio de los resultados obtenidos en el paso anterior y, como segundo umbral, valor máximo alcanzado por los mismos.
- Las funciones que delimitan las bandas se consiguen al restar y sumar a la serie observada dichos umbrales y, posteriormente, retrasar y adelantar el resultado 1 hora.

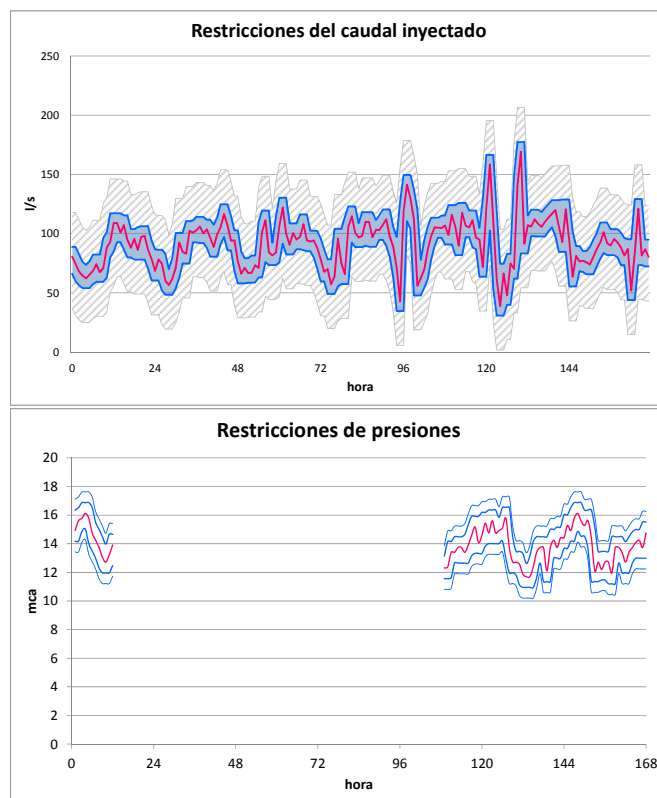


Figura 37: Ejemplo de los valores medios y umbrales admisibles para el caudal inyectado y las presiones, necesarios para el cálculo de la función objetivo.

En el caso del único punto de caudal inyectado medido, se dispone de varios días solapados para obtener los resultados descritos. Sin embargo, en el caso de las presiones, de ocho puntos disponibles para el proceso de calibración, únicamente dos, P2 y P6, cuentan con registros coincidentes en alguna franja horaria. La decisión tomada fue imponer al resto de registros los umbrales más laxos entre los calculados a partir de los datos asociados a estos dos puntos (ver Tabla 28).

	1 ^{er} umbral	2 ^o umbral
Caudal inyectado (l/s)	8.22	37.24
Presiones (mca)	0.75	1.5

Tabla 28: Valores de los umbrales para el caudal inyectado y las presiones observadas.

3.6.2. Variables de decisión

Las variables de decisión consideradas en el caso de estudio son:

- Rugosidad de las tuberías
- Coeficientes de descarga
- Factores de demanda
- Demanda base de los nudos de consumo

La inclusión de esta última en la calibración, provoca que el número total de las mismas aumente considerablemente. Ésta es una de las razones que justifica el clústering de tuberías, referente a su rugosidad, y el clústering de nudos, asociado a los coeficientes de descarga, pues con dichos análisis se alcanza el propósito de reducir la cantidad de variables de decisión. Estos dos clústering se desarrollan en el apartado 3.5.5.

<u>Variables de decisión</u>	<u>Número total</u>
Rugosidades de las tuberías	25
Coeficientes de descarga	39
Factores de demanda	168
Demanda base de los nudos de consumo	160

Tabla 29: Número total de elementos por cada tipo de variable de decisión.

La **Tabla 29** especifica el número de elementos considerado en cada uno de los cuatro tipos de variables de decisión, ascendiendo el número total a 392.

3.6.2.1. Valores extremos de los distintos tipos de variables de decisión

El espacio en el que busca el algoritmo depende de los rangos asociados a las variables de decisión, dentro de los cuales se le permite crear nuevos individuos o, dicho de otro modo, de las restricciones impuestas a las variables de decisión. Para establecer dichas restricciones, se llevaron a cabo varias pruebas. En primer lugar, se tuvieron en consideración rangos amplios para las *rugosidades* (0.01-1) y los *coeficientes de descarga* (0.0001-2), observando que éstos no eran adecuados, pues el algoritmo encontraba resultados inverosímiles: se asignaba a las fugas el 80% del volumen inyectado y existían presiones negativas en el sistema. Seguidamente, se probaron rangos drásticamente más acotados y las soluciones pivotaron al otro extremo: fugas en torno al 5% del volumen inyectado. A partir de éstos, los límites máximos se fueron aumentando progresivamente hasta conseguir resultados comprensibles, relacionados con valores de las variables de decisión que en general no se ubicaban cerca de los límites. En el caso de los *factores de demanda*, también se empezó por probar un rango amplio, pero no fue necesario corregirlo. Con dicho conjunto de pruebas, se fijaron los límites de las tres variables citadas, mostrados en la **Tabla 30**.

Mínimo Máximo

Rugosidades de las tuberías	0.01	0.35
Coefficientes de descarga	0.01	0.2
Factores de demanda	0	2.2

Tabla 30: Rangos mínimos y máximos para las variables de decisión (excepto la *demanda base*).

Los rangos para la *demanda base* de los nudos de consumo requieren otro tipo de procedimiento. El modelo definido para el caso de estudio consta de 160 nudos de consumo. Asociar los mismos rangos a todos ellos no sería adecuado, pues se posee información que puede restringir más el espacio de búsqueda y que, en principio, haría que el ajuste fuera más fiel a la realidad. Para fijar los valores extremos se utilizó la muestra aleatoria conseguida según los pasos explicados en el apartado 3.5.3.1.3. Dichos valores extremos se asociaron al valor mínimo y al máximo de la muestra correspondiente, para cada uno de los primeros 30 grupos determinados en el clústering de nudos de consumo. Los grupos 31 y 32 son un caso especial, dado que se crearon ad hoc para dos nudos a los que se adjudicaron parcelas ubicadas fuera de la planta urbana consolidada, de las cuales no se conoce ni sus características ni el número exacto al que ascienden. Por este motivo, y únicamente para dichos grupo, se establecen sus límites en el valor mínimo y máximo definido por el resto de grupos. En la **Tabla 31** se especifican todos estos umbrales.

Grupo	Mínimo	Máximo
1	0.1206	0.3426
2	0.0961	0.3439
3	0.0793	0.2546
4	0.0787	0.2892
5	0.4286	0.7809
6	0.1860	0.4078
7	0.3374	0.6576
8	0.2490	0.5336
9	0.1105	0.3054
10	0.3376	0.6225
11	0.2272	0.4850
12	0.1539	0.4240
13	0.5372	0.9039
14	0.2531	0.4979
15	0.5534	0.9448
16	0.2795	0.5633
17	0.3771	0.7529
18	0.5112	0.9361
19	0.2818	0.5959
20	0.1647	0.4703
21	0.2903	0.6015
22	0.4171	0.7459
23	0.3816	0.6966
24	0.6865	1.0764
25	0.9892	1.4664
26	0.1815	0.4381
27	0.1877	0.4398

28	0.0924	0.2745
29	0.4809	0.8032
30	0.4199	0.7942
31	0.0787	1.4664
32	0.0787	1.4664

Tabla 31: Valores mínimos y máximos para las demandas base de nudos de consumo, diferenciados por clúster.

3.6.3. Función objetivo

Un aspecto importante en los procesos de calibración es la función objetivo que se asocia al problema. Ésta es la que va a guiar al algoritmo hacia las mejores soluciones, en consecuencia, es importante definirla con criterio y adecuadamente.

Como este estudio se alimenta de información con un grado de incertidumbre no despreciable, se emplea el concepto de función de pertenencia de los subconjunto difuso para definir la función objetivo. El propósito es no descartar indiscriminadamente todas las soluciones que no emulen con un grado de precisión alto los datos observados.

Llegados a este punto, se plantea la siguiente función objetivo:

$$FO = P + Q + DBp + DBc + RD$$

El valor al que debe tender dicha función objetivo es 0. Los cinco términos que componen la ecuación anterior son:

- ***P*** : Presiones. Valor máximo admisible: 8.
- ***Q*** : Caudal inyectado. Valor máximo admisible: 2 (esto implica que las soluciones adoptadas pueden ubicar algún punto de la serie de caudal inyectado desde el tanque nro. 35 más allá del segundo umbral).
- ***DBp*** : Término relacionado con la distribución de la población (asociada también, aunque indirectamente, a la demanda base). Valor máximo admisible: 1.
- ***DBc*** : Demanda base de los nudos de consumo. Valor máximo admisible: 1.
- ***RD*** : Restricción dura.

Durante el desarrollo de este trabajo se ha analizado otra función objetivo. En ella, se proporcionaban distintos pesos a cada uno de los términos citados previamente. De este modo, se pretendía dar un poco más de importancia a la modulación del caudal inyectado frente a la presiones. Las pruebas llevadas a cabo demostraron que la función con los términos ponderados no conseguía el propósito para el que fue concebida. A su vez, con la misma se obtenían peores resultados en las presiones. A causa de esto, dicha función objetivo fue descartada a favor de la descrita en el presente apartado.

3.6.3.1. Términos asociados con observaciones de caudal inyectado y presiones

Los dos primeros términos de la función objetivo están relacionados con registros observados, por lo que el valor que tomen dependerá de cómo se parezcan éstos a los simulados. Normalmente, se emplea el error cuadrático medio como medida de la precisión del ajuste. Sin embargo, como la función objetivo incluye variables de distinta magnitud, el empleo de esta medida del error conllevaría que los términos de la función objetivo también se encuentren en órdenes de magnitud diferente. Para solventar esto, dentro del sistema de penalizaciones que se define a continuación, se emplea un cálculo similar al realizado en la ecuación de Nash-Sutcliffe (Nash, 1970), el cual permite normalizar los valores de penalización y hacer comparables los distintos términos de la función objetivo.

3.6.3.1.1. Presiones

Para cada uno de los ocho puntos con registros de presión observada que se emplean en esta calibración, se calcula un valor de penalización (PP_i):

$$PP_i = \text{Promedio}(PP_{i-h})$$

donde $i=1, 2, \dots, 8$; h es la hora=1, 2, ..., 168 y la penalización PP_{i-h} se obtiene según la función mostrada en la **Figura 38**.

El empleo del *promedio* en vez de la *suma* en la ecuación anterior es debido a que, para cada punto de presión, existe un número distinto de registros con los cuales evaluar la penalización. Al emplear la *suma*, tendría más importancia en la función objetivo aquellos puntos con registros más amplios.

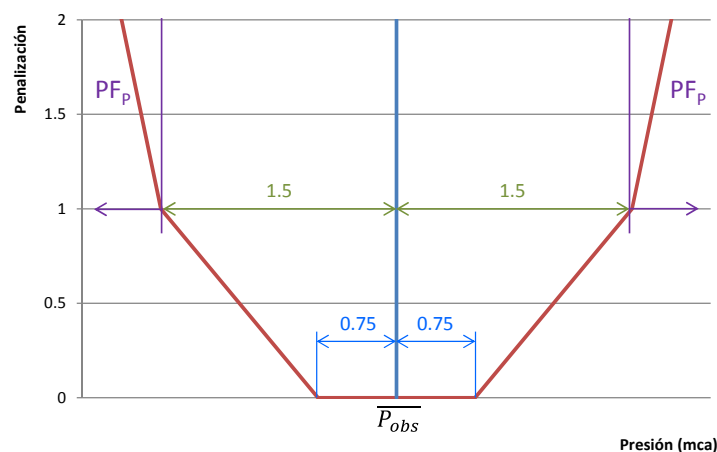


Figura 38: Cálculo de la penalización PP_{i-h} asociada a las presiones, evaluada para cada hora.

La **Figura 38** distingue tres zonas de penalización definidas a partir del valor medio de la presión observada. La primera zona abarca 0.75 mca a ambos lados de dicho valor medio y la penalización asociada es nula. La segunda llega hasta los 1.5 mca y define el umbral máximo admisible, hasta el cual la penalización crece linealmente de 0 a 1, iniciándose ésta

allí donde acaba el primer umbral. Por último, la tercera comprende el resto de valores de presión y se evalúa mediante el término PF_p . En el apartado 3.6.1.2 se detalla cómo se han obtenido los valores que describen estas zonas.

$$PF_p = \frac{(P_{i-h_{sim}} - P_{i-h_{obs}})^2}{(P_{i-h_{obs}} - \overline{P_{obs}})^2} * 100$$

Como se ha comentado y se aprecia en la ecuación anterior, PF_p se estima basándose en el índice de Nash-Sutcliffe. En este caso particular no se calcula para la serie completa, sino para cada paso de tiempo. A su vez, el resultado de esa normalización se multiplica por un factor, permitiendo descartar con mayor facilidad las soluciones que están fuera del umbral máximo. Este hecho viene ocasionado por el empleo del *promedio* en el cálculo de PP_i lo que enmascara la penalización dada en una hora concreta. El valor empleado para dicho factor es 100, escogido tras probar distintas opciones.

Con todo esto, el valor final del término de presiones será admisible, en principio, cuando esté por debajo de 8:

$$P = \sum_{i=0}^8 PP_i$$

3.6.3.1.2. Caudal inyectado

El término asociado al caudal inyectado se evalúa de forma similar a lo explicado para las presiones. La penalización PQ_{iny} se evalúa como:

$$PQ_{iny} = Promedio(PQ_{iny-h})$$

donde h es la hora=1, 2,..., 168 y PQ_{iny-h} se obtiene a partir de la **Figura 39**.

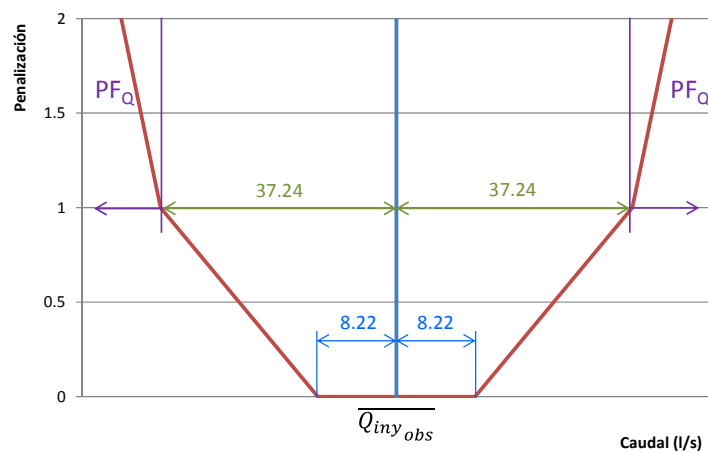


Figura 39: Cálculo de la penalización PQ_{iny-h} asociada a los caudales inyectados, evaluada para cada hora.

El término PF_Q para el caudal inyectado se calcula mediante la siguiente ecuación:

$$PF_Q = \frac{(Q_{iny-h_{sim}} - Q_{iny-h_{obs}})^2}{(Q_{iny-h_{obs}} - Q_{iny_{obs}})^2} * 100$$

Con todo esto, el valor final del término Q de la función objetivo equivale a PQ_{iny} y se considerará admisible por debajo de 2:

$$Q = PQ_{iny}$$

3.6.3.2. Términos asociados con la demanda base de consumo

3.6.3.2.1. Población

Este término pretende asegurar que la distribución de la demanda base sea coherente con la cantidad de población a la que abastece. Para ello se delimitan 3 zonas en una capa SIG en las que se solapan los radios censales (*802_RadioCensalCenso2010*), que incluyen los datos de población, y las áreas de influencia asociadas a los nudos de consumo (*902_AreasInfluenciaMallas1*). En dichas zonas son conocidos, en consecuencia, el caudal total abastecido y la cantidad de población que previsiblemente lo consume.

La función de penalización definida en este caso es de tipo triangular:

- En el vértice la penalización toma el valor cero. En dicho punto, la proporción de población en una zona con respecto a la población total es igual a la proporción del caudal consumido en dicha zona respecto del total demandado por los usuarios.
- Conforme la proporción de la población en una zona dista de la proporción de caudal consumido, la penalización va aumentando linealmente, hasta alcanzar un valor de 1 cuando dicha diferencia es del 10%. A partir de ésta, si sigue aumentando los resultados se consideran no admisibles y la penalización pasa a ser fuerte, multiplicando por un valor de 1000 la diferencia entre la proporción correspondiente al vértice y la obtenida en la iteración evaluada.

El promedio de la penalización obtenida en cada una de las zonas da como resultado DBp y se considera admisible por debajo de 1.

3.6.3.2.2. Demanda base en los nudos de consumo

La demanda base de cada grupo de nudos de consumo está caracterizada por una función de densidad (apartado 3.5.3.1.2). En un intento de guiar al algoritmo genético hacia valores de esta variable más próximos al medio definido por dicha función de densidad, se crea una función de penalización trapezoidal que:

- Toma el valor cero mientras la demanda base de cualquier nudo j , perteneciente al grupo i , quede comprendida en el rango establecido por el área bajo la curva de la función de densidad correspondiente al 70% de la misma.
- A partir de ese rango, la penalización va aumentando linealmente hasta 1 conforme se aproxima a los límites máximos.

El promedio de las penalizaciones generadas por cada nudo de consumo es el valor asociado al término DBc y se considera admisible por debajo de 1.

3.6.3.3. Restricción dura

El último término de la función objetivo se introduce para descartar aquellas soluciones que, una vez generada la nueva configuración por el algoritmo, no cumplen un criterio básico:

$$\overline{Q_{cons}} < 90\% \overline{Q_{iny_T35}}$$

Es decir, el caudal medio consumido debe ser menor que el 90% del caudal medio inyectado en el tanque principal nro. 35. Recordar que dicho tanque es el que proporciona aproximadamente el 85% del caudal inyectado a red. Si este criterio se cumple, el valor de RD es 0, en caso contrario:

$$RD = (\overline{Q_{cons}} - 90\% \overline{Q_{iny_T35}}) * 1000$$

3.6.4. Pruebas realizadas

Evolver emplea un algoritmo genético que funciona según unos parámetros de entrada, entre los que destacan: el tamaño de la población, la tasa de cruce, la tasa de mutación y la semilla. Los valores que se les asigne a éstos, afectarán de un modo u otro en la búsqueda de soluciones. Los tres primeros condicionarán la evolución de la población, mientras que la semilla determinará la generación de los números aleatorios presentes en todo el proceso. Es por esto que la elección de la combinación de estos parámetros es significativa en la obtención de resultados en el proceso de calibración del modelo.

Debido a lo comentado en el párrafo anterior, a falta de medios computacionales, frente al desconocimiento sobre qué combinación es la más idónea y, sobre todo, teniendo en cuenta que se buscan distintas configuraciones del sistema que proporcionen unos resultados dentro de los rangos admisibles, se propone el lanzamiento de una serie de pruebas en las que los parámetros del algoritmo son elegidos aleatoriamente entre un conjunto concreto de valores definido a partir de los recomendados por el manual de Evolver (ver **Tabla 32**).

Parámetro	Valores posibles
Tamaño de la población	50; 75; 100; 125; 150
Tasa de cruce	0.4; 0.45; 0.5; 0.55; 0.6
Tasa de mutación	0.05; 0.1; 0.15
Semilla del nº aleatorio	Automática. Distinta para cada calibración.

Tabla 32: Valores posibles de los parámetros del algoritmo genético.

Otro aspecto a tener en cuenta cuando se trabaja con algoritmos genéticos es definir el criterio de parada del mismo. Evolver permite elegir varios simultáneamente:

- Número máximo de evaluaciones
- Tiempo máximo
- Progreso
- Fórmula

Los dos primeros permiten fácilmente decidir cuánto tiempo tiene el algoritmo para encontrar una posible solución, aunque éste depende del ordenador en el que se lleve a término el proceso de calibración. En el criterio de progreso se concretan dos parámetros: porcentaje (%P) y número de iteraciones (NI). El objetivo de dicho criterio es finalizar el proceso cuando durante el NI fijado, el valor de la función objetivo no varía en el %P establecido. El último criterio se basa en elegir una fórmula tal que en el momento se cumplen los objetivos de la calibración finaliza el proceso.

Para el presente caso de estudio, el criterio de progreso es más adecuado que el resto, puesto que es el menos estricto y depende del resultado que se va obteniendo. Sin embargo, en previsión de que ese criterio no finalice en un tiempo razonable y teniendo en cuenta que los recursos disponibles son limitados, se emplea también el número máximo de iteraciones. Los valores establecidos para cada criterio se encuentran en la **Tabla 33**. Éstos son bastante laxos desde el punto de vista de finalización, con lo cual se dota al algoritmo de un margen de maniobra amplio en este sentido, en tanto que tiene que probar una gran cantidad de combinaciones de las variables de decisión.

Criterio	Valor
Número máximo de evaluaciones	150000
%P	0.01%
NI	10000

Tabla 33: Valores establecidos para los criterios de parada del algoritmo genético.

Teniendo en cuenta todo lo descrito, en este apartado de calibración se procedió, con el tiempo disponible, a lanzar las calibraciones mostradas en la **Tabla 34**.

Prueba	Semilla	Tamaño población	Tasa de cruce	Tasa de mutación	Lím. Máx. Rugosidad	Lím. Máx. Coef. descarga
A1	162592248	125	0.4	0.15	0.35	0.2
B1	301889873	75	0.55	0.15	0.35	0.2
C1	958881702	50	0.6	0.1	0.35	0.2
D1	192410576	100	0.4	0.05	0.35	0.2
E1	100012295	150	0.6	0.1	0.35	0.2
A2	188025998	50	0.6	0.15	0.25	0.1
B2	330068107	75	0.4	0.05	0.25	0.1
C2	970176206	150	0.5	0.1	0.25	0.1
D2	286838795	50	0.45	0.1	0.25	0.1
E2	117366498	100	0.6	0.15	0.25	0.1
D3	325435841	50	0.6	0.05	0.35	0.1
B3	379664826	75	0.45	0.15	0.35	0.1
E5	45590216	125	0.4	0.05	0.35	0.1
A5	250108873	125	0.45	0.15	0.35	0.1
E6	62258466	125	0.6	0.05	0.35	0.1
E3	644482	75	0.4	0.15	0.25	0.2
C3	997813671	125	0.45	0.05	0.25	0.2
A3	205325263	125	0.5	0.15	0.25	0.2
E4	37010091	50	0.4	0.15	0.25	0.2
C4	1029598188	125	0.4	0.1	0.25	0.2
A4	234254185	75	0.6	0.1	0.35	0.2
C5	1055100420	125	0.6	0.1	0.35	0.2
E7	78082670	125	0.5	0.05	0.35	0.2
A6	277013841	50	0.4	0.05	0.35	0.2

Tabla 34: Calibraciones realizadas para la obtención de diferentes configuraciones admisibles del sistema de suministro de agua de Nogoyá.

El conjunto de lanzamientos se ha llevado a su fin mediante el empleo de cuatro ordenadores personales. Atendiendo únicamente al primer criterio de parada (número máximo de evaluaciones), el tiempo aproximado de simulación es de 12 horas para dos de ellos y de 7 horas para los otros dos.

3.7. Análisis de resultados

Las pruebas que finalmente han alcanzado valores de la función objetivo admisibles pueden consultarse en la **Tabla 35**. El resto se encuentran en el Anejo 4.

Prueba	Semilla	Tamaño población	Tasa de cruce	Tasa de mutación	Lím. Máx. Rugosidad	Lím. Máx. Coef. descarga	FO	P	Q	DBp	DBc	Criterio de parada
C1	958881702	50	0.6	0.1	0.35	0.2	2.07	0.49	1.33	0.16	0.09	P
D1	192410576	100	0.4	0.05	0.35	0.2	4.48	2.39	1.66	0.33	0.1	P
E1	100012295	150	0.6	0.1	0.35	0.2	2.67	0.66	1.55	0.24	0.22	NME
C3	330068107	75	0.4	0.05	0.25	0.1	7.06	4.98	1.62	0.22	0.23	NME
B2	45590216	125	0.4	0.05	0.35	0.1	2.79	0.7	1.55	0.25	0.29	P
E5	62258466	125	0.6	0.05	0.35	0.1	2.44	0.49	1.5	0.27	0.17	P
E6	997813671	125	0.45	0.05	0.25	0.2	3.62	1.74	1.55	0.25	0.08	NME
E7	78082670	125	0.5	0.05	0.35	0.2	1.95	0.45	1.22	0.23	0.03	NME
A6	277013841	50	0.4	0.05	0.35	0.2	3.54	1.59	1.6	0.22	0.13	P

Tabla 35: Para las calibraciones lanzadas que obtienen resultados admisibles, valores alcanzados por la función objetivo y por cada uno de los término que forman parte de ésta, siendo P=Progreso y NME=Número máximo de evaluaciones.

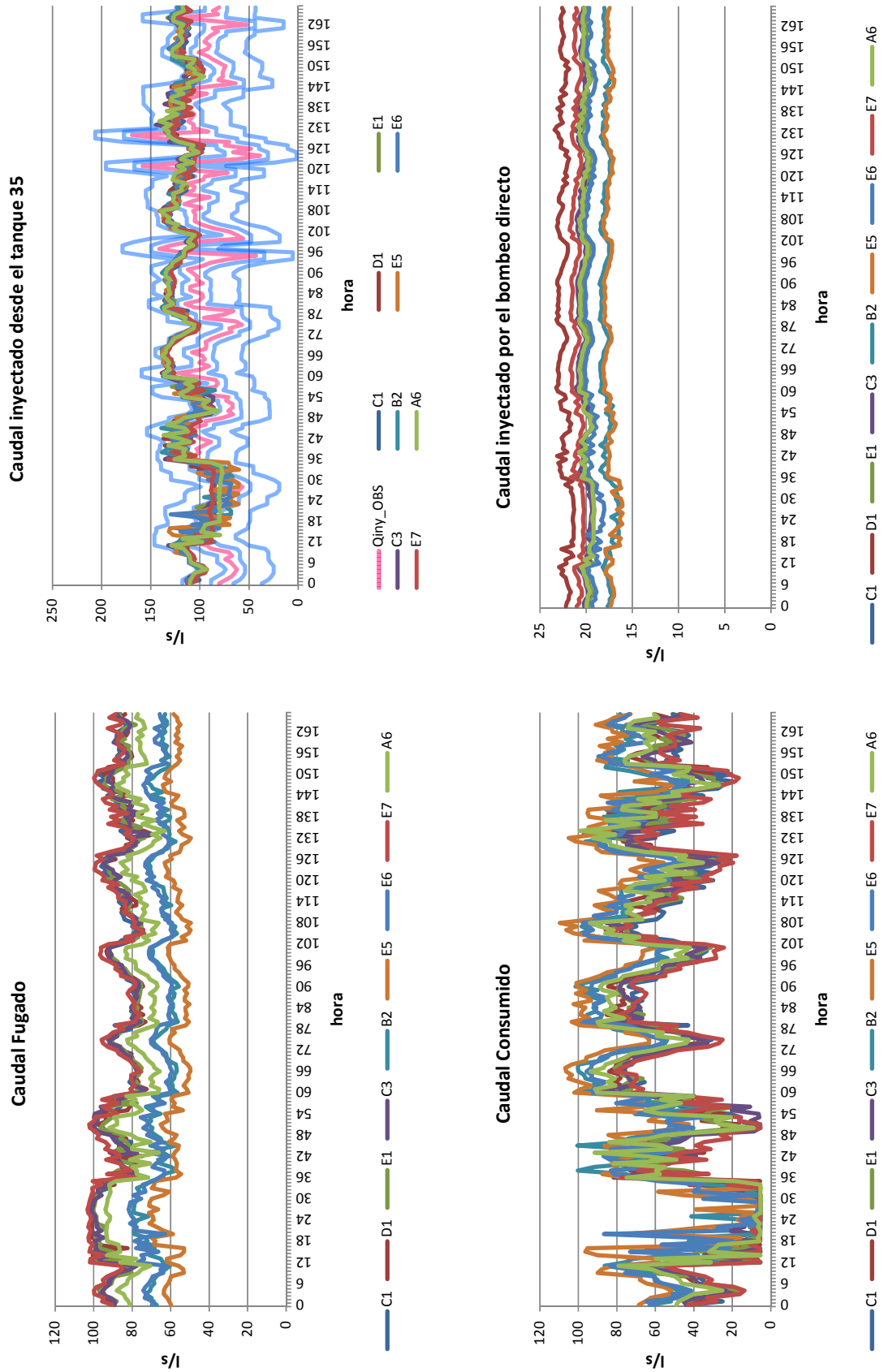
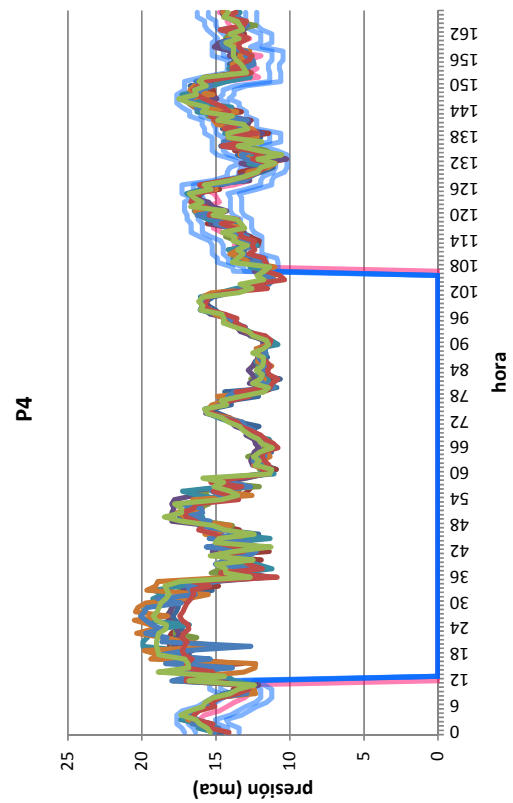
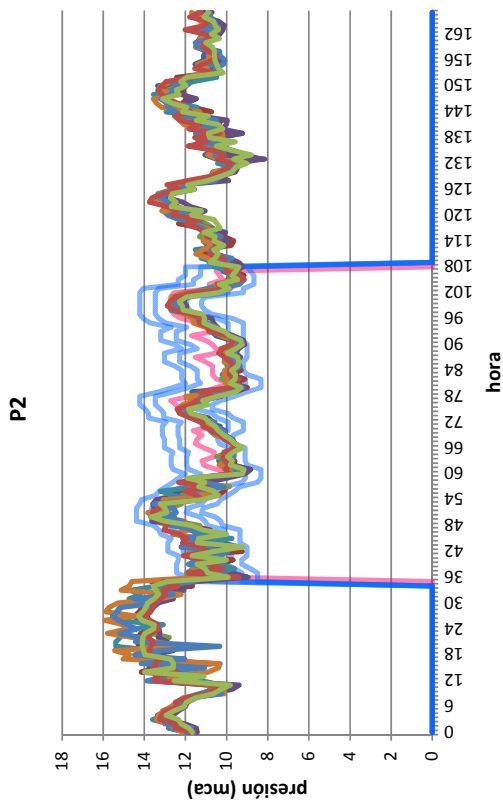
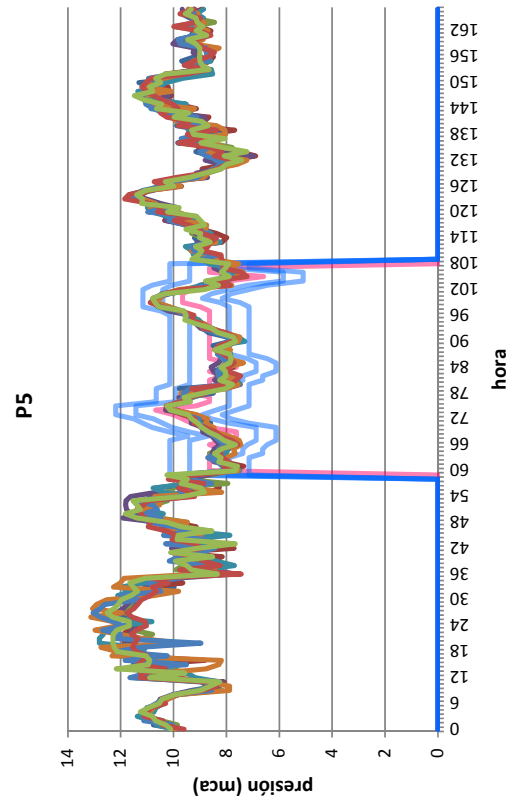
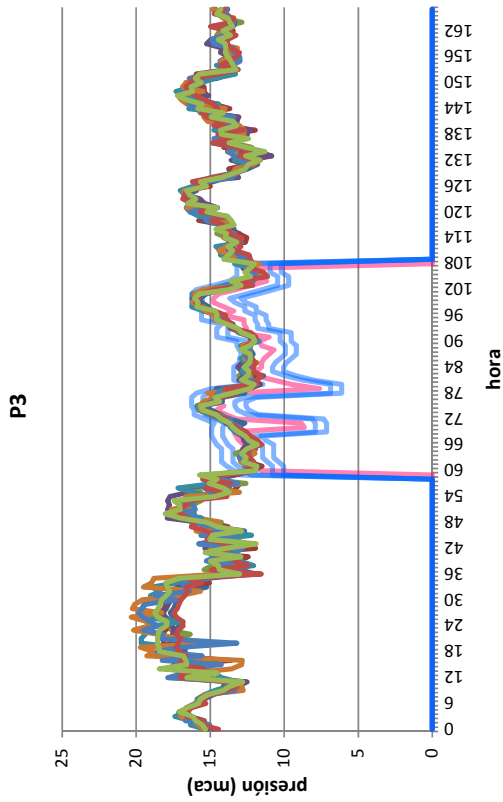


Figura 40: Resultados de las calibraciones consideradas como admisibles en referencia al caudal inyectado, consumido y fugado.

Los resultados obtenidos por dichas pruebas referentes al caudal *inyectado*, *consumido* y *fugado* quedan resumidos en la **Figura 40**. Las presiones simuladas en los puntos allí donde existían mediciones sobre esta variable pueden consultarse en la **Figura 41**.

Las conclusiones que se extraen a primera vista son:

- Las distintas variables presentan mayor rango de variación el martes, único día en el que no existen mediciones de presión. A éste le siguen en poca cantidad de registros disponibles el sábado, domingo y lunes, lo que también se aprecia en la banda definida de posibles soluciones. Precisamente gracias a esto, los martes es el día que el caudal inyectado se parece más al observado. Llama poderosamente la atención que esto es a costa de que el caudal fugado aumente y el consumido (tratado como independiente de la presión) disminuya drásticamente, así como, de forma coherente, que en todos los puntos con presiones observadas aumente las mismas con respecto a otros días.
- Por otro lado, queda patente que el volumen *fugado*, *inyectado mediante bombeo directo* y *consumido* se acomodan de distinto modo para dar como resultado el caudal inyectado desde el tanque nro. 35. Éste constituye otro ejemplo más de como las variables desconocidas son las que absorben su propia incertidumbre en los resultados, mostrando un rango de variación con una amplitud considerablemente mayor al de aquellas que están más acotadas.
- En ninguna de las simulaciones llevadas a cabo ha sido posible encontrar una configuración del sistema que situara el volumen fugado por debajo del 50% del volumen total inyectado.
- La modulación del caudal inyectado desde el tanque nro. 35 no ha conseguido reproducirse correctamente, sobre todo los dos picos nocturnos que tienen lugar jueves y viernes a la media noche. Además, todas las soluciones tienden a mayorar dicho caudal. En las conclusiones del estudio se hará referencia a este hecho.
- El caudal consumido es el que muestra mayor fluctuación. Esto es previsiblemente así debido a que dos variables de decisión, que pueden ser de distinta manera combinadas entre sí, son las que afectan a este resultado.



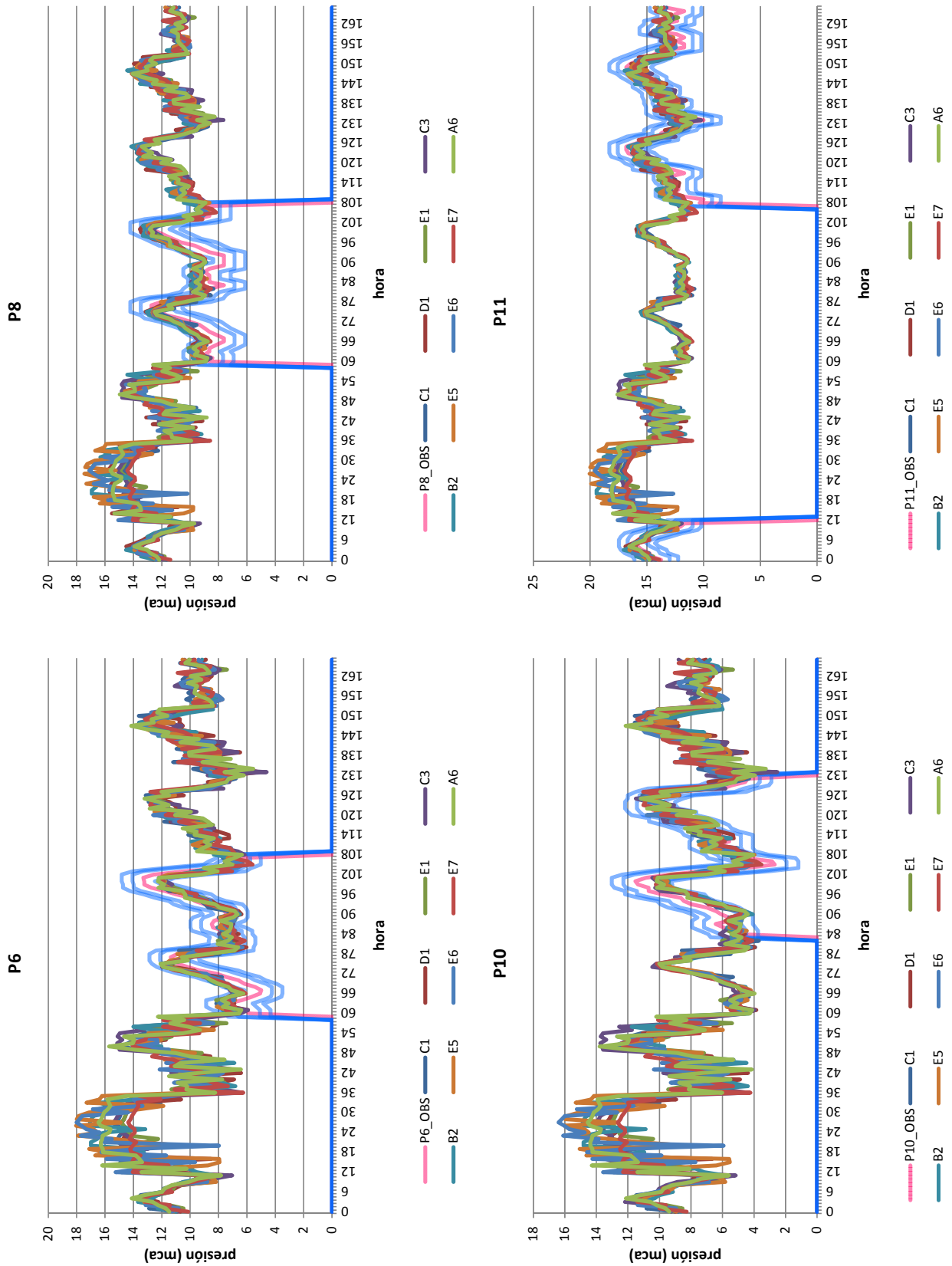


Figura 41: Presiones simuladas en los puntos de la red allí donde se habían llevado a cabo observaciones de la misma.

Los rangos dentro de los cuales se encuentran los valores tomados por las variables de decisión se analizan a continuación. En primer lugar, se observa con respecto a la *rugosidad* que aquellas tuberías pertenecientes a clústeres cuyo rango de variación posee un intervalo intercuartílico menor que 0.1 (**Figura 43**; dichos clústeres en concreto son 4, 5, 6, 8, 10, 12, 15, 16, 17, 18, 19, 20, 21, 25) conducen agua hasta zonas que abastecen a mayor número de parcelas, así como zonas en las que se ha cargado mayor cantidad de fugas en los distintos modelos (**Figura 42**). Una posible explicación es que al tener mayor incidencia en las presiones de las distintas configuraciones del sistema encontradas, no todos los posibles valores del rango tienen cabida a la hora de emular los datos empleados para la calibración. Las agrupaciones que han alcanzado una mediana por encima de 0.2 (**Figura 42**) son de distinto diámetro, material y edad.

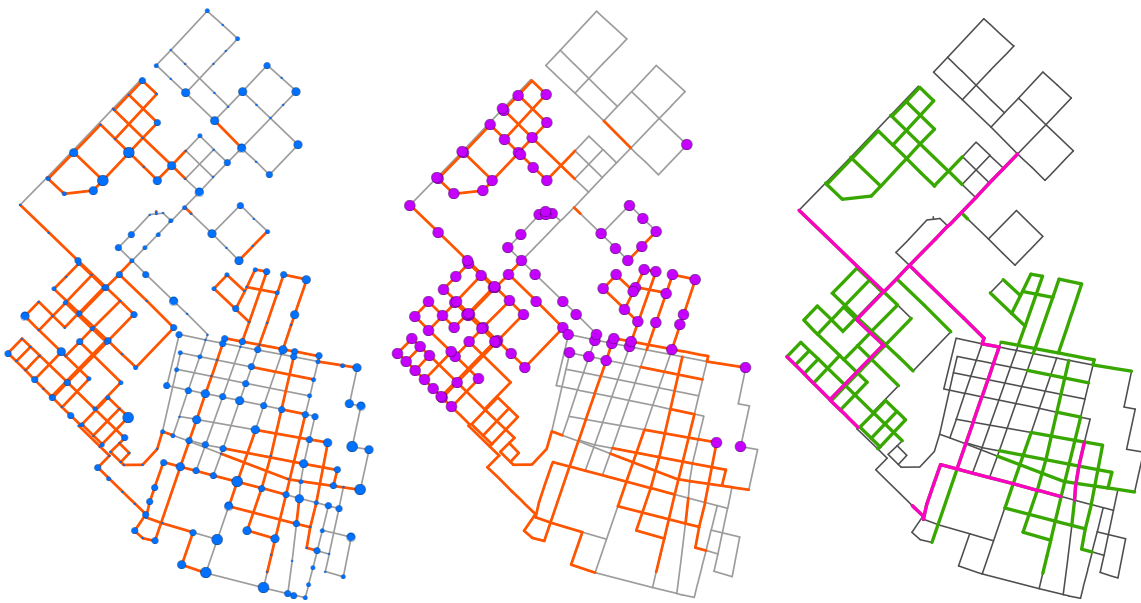


Figura 42: Las tuberías resaltadas en **naranja** corresponden a aquellos grupos con un rango de variación en la *rugosidad* más estrecho con respecto al resto. En la figura de la izquierda, los puntos **azules** tienen un tamaño proporcional al número de parcelas totales que abastecen cada nudo (los más pequeños corresponden a cero, es decir, únicamente son nudos de conexión). En la figura del centro, los puntos **morados** son los asociados a aquellos grupos con coeficientes de descarga cuya media es más alta en relación a la del resto. Por último, en la figura de la izquierda se destacan las tuberías cuya mediana está por encima de 0.2, siendo las de color **fucsia** las que corresponden a diámetros superiores a 100 mm.

Por otro lado, los grupos a los que se les ha adjudicado un *coeficiente de descarga* con una mediana por encima de 0.2 (**Figura 43**) no corresponden a aquella zona con mayor número de acometidas y longitud de tubería ponderada por nudo (**Figura 42**). Los clúster con un intervalo intercuartílico por debajo de 0.05 tienden normalmente a tener medianas bajas (**Figura 42**). La lectura general parece indicar que existe un área a la que se le ha asignado un volumen importante de fugas y, en consecuencia, ello conlleva a que queden acotados los posibles valores que toman un amplio número de puntos en otras zonas.

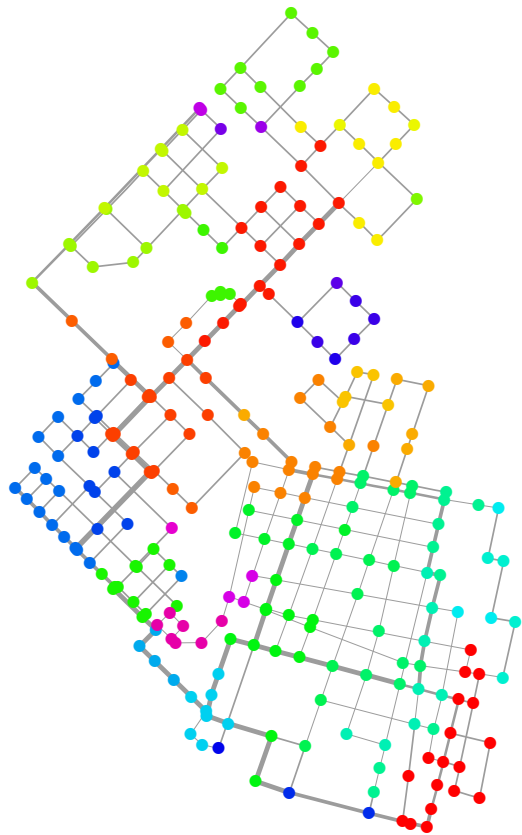
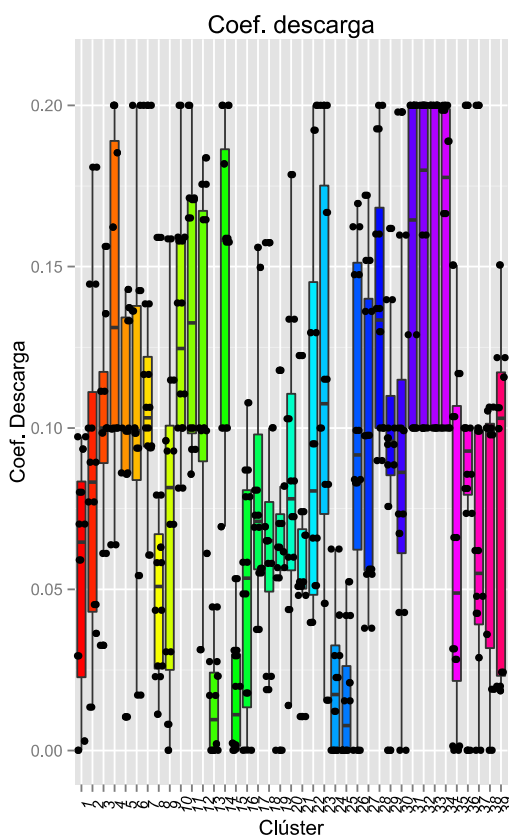
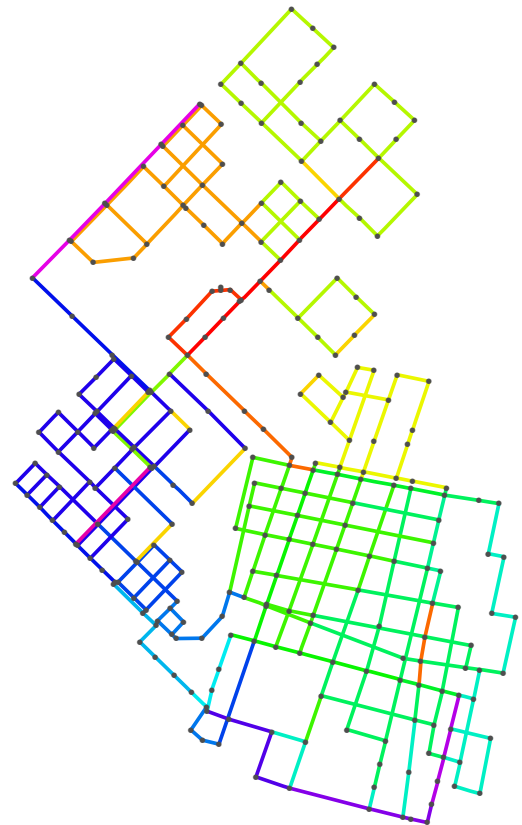
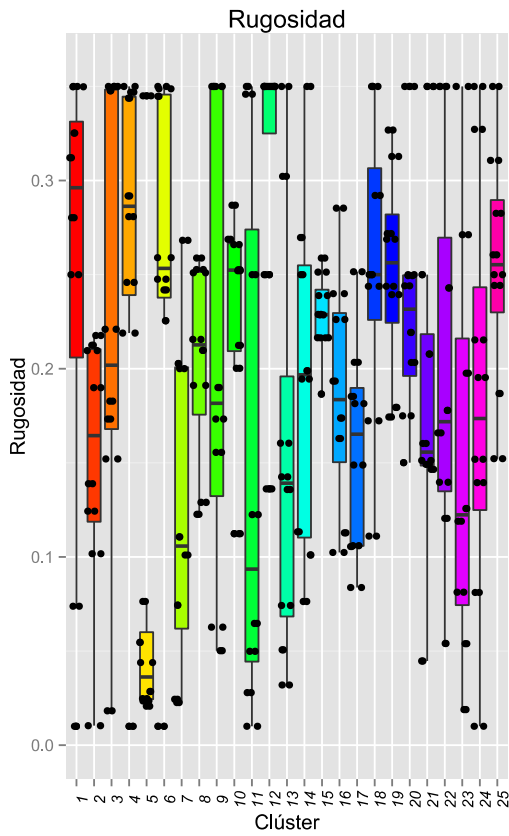


Figura 43: Resultados obtenidos para las variables de decisión rugosidad y coeficiente de descarga.



Figura 44: Los puntos **morados** son los asociados a aquellos grupos con coeficientes de descarga cuya media es más alta en relación a la del resto. Los puntos **verdes** tienen un tamaño proporcional a la longitud ponderada de tubería y los **azules** al número de acometidas, ambos valores estimados para el clústering de coeficientes de descarga. La zona con un número de acometidas y longitud ponderada de tubería en media alto está rodeada por una circunferencia. En la figura de la izquierda se resaltan los nudos en **naranja** que forman parte de clústeres con intervalos intercuartílicos menores a 0.05.

Para finalizar el análisis de las variables de decisión, en la **Figura 46** y **Figura 46** aparecen resumidos los resultados obtenidos con respecto a la *demanda base* y los *factores de demanda*. A primera vista resalta la amplia variabilidad en ambas, tal y como ya se preveía al analizar el caudal consumido. Con respecto a este último, también resulta coherente que los *factores de demanda* del martes estén prácticamente sobre el límite inferior. En general, la *demanda base* atribuida a los distintos nudos de consumo se mueve por todo el rango permitido, no centrándose específicamente en el más estricto donde la penalización era cero.

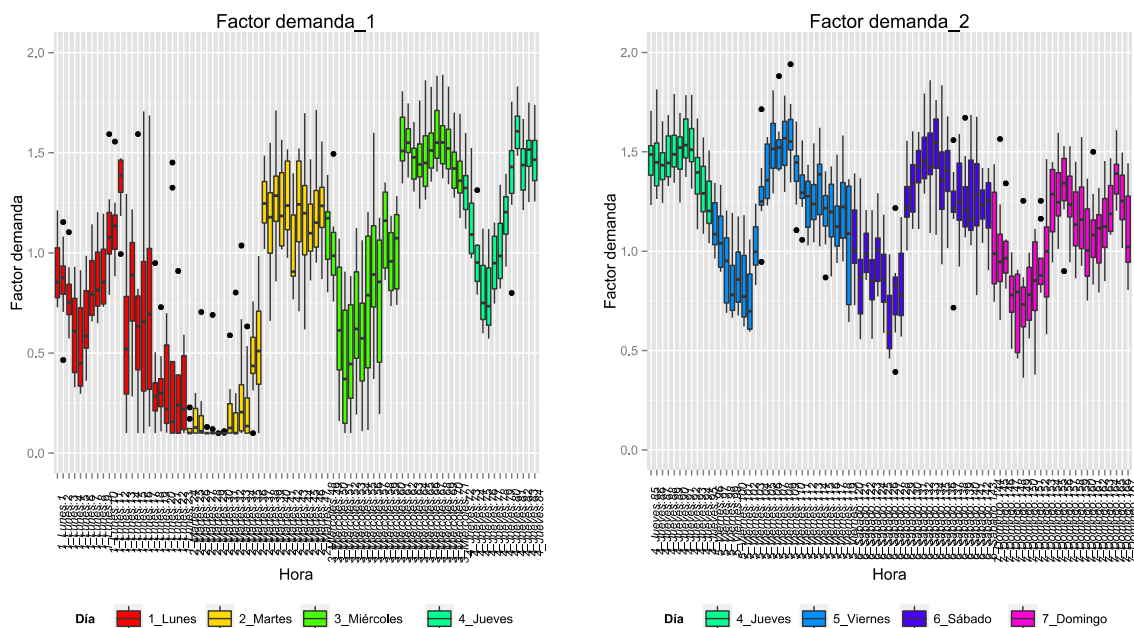


Figura 45: Resultados obtenidos para la variable de decisión *factores de demanda*.

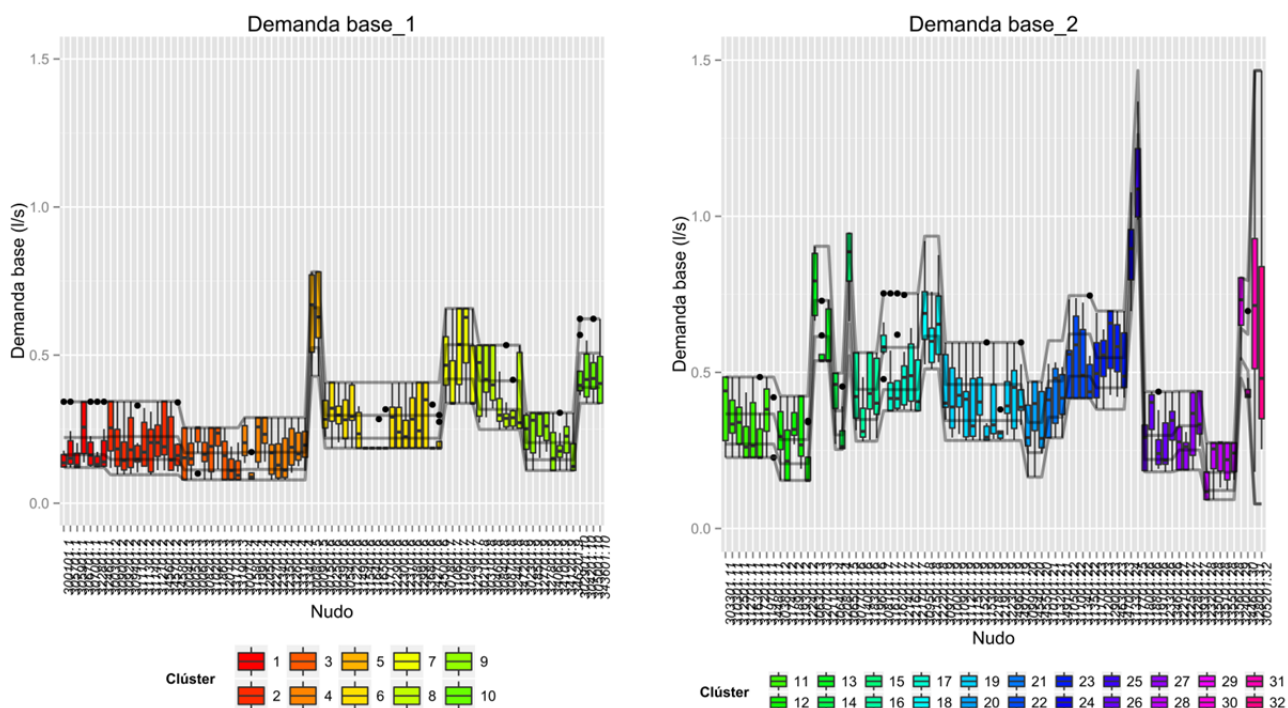
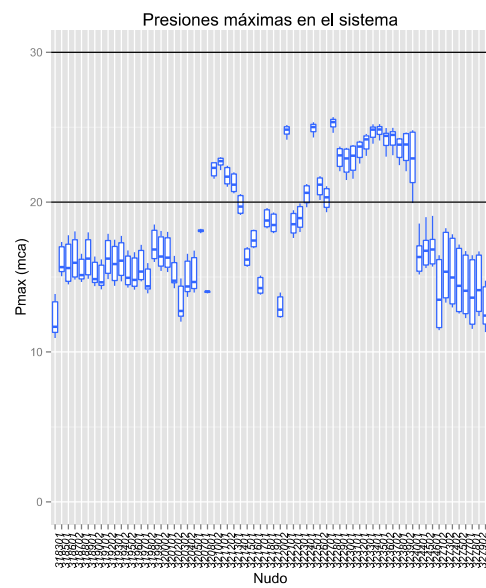
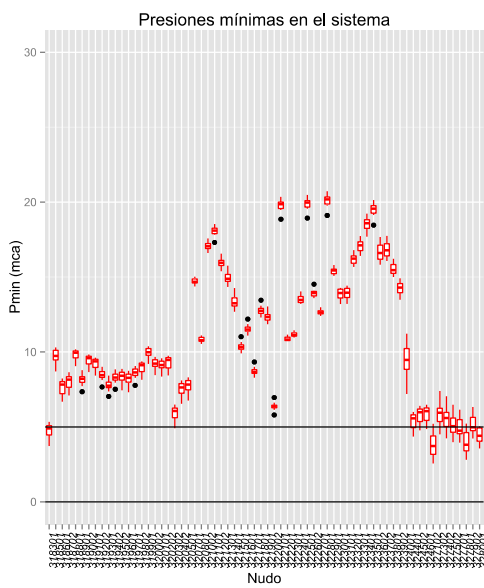
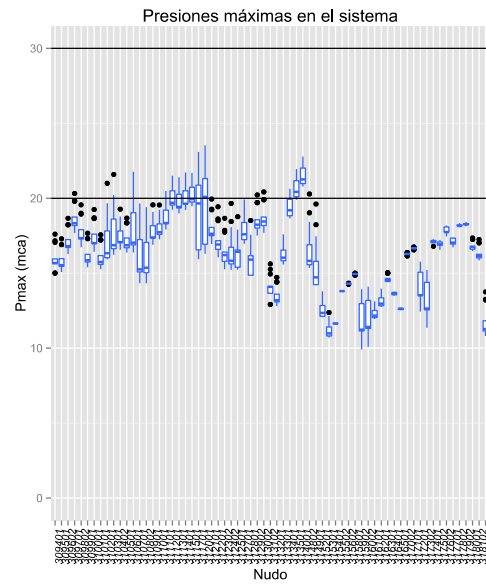
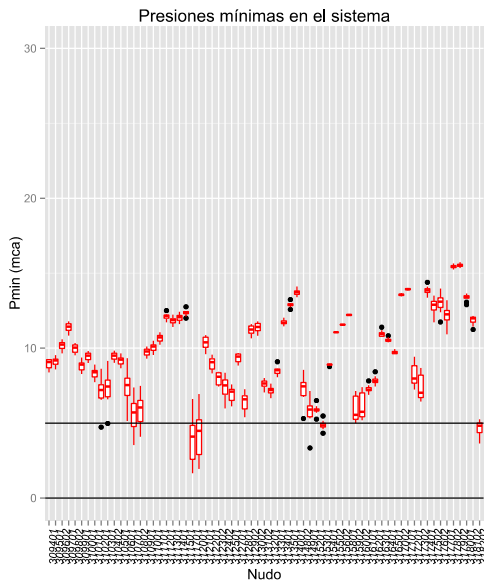
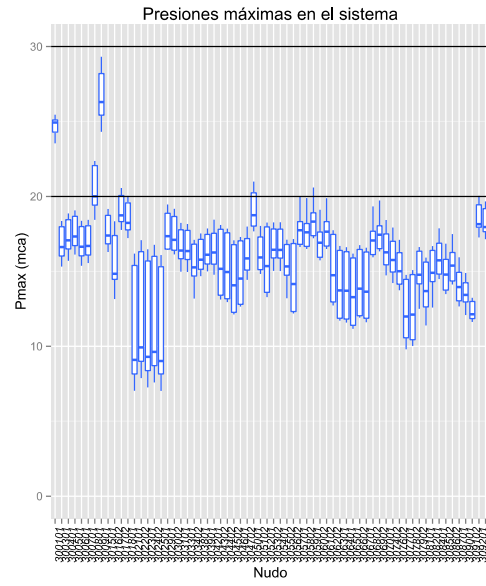
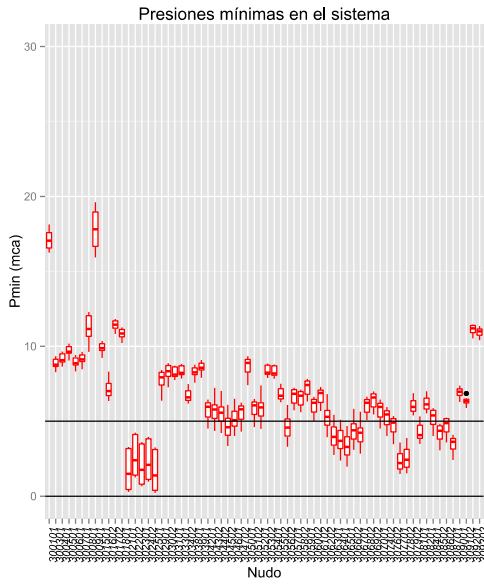


Figura 46: Resultados obtenidos para la variable de decisión *demanda base* de los nudos de consumo.

Seguidamente, se procede a comentar los resultados obtenidos en relación a las presiones alcanzadas en el sistema, ya que cabe recordar que conocer las mismas frente a un incremento de la demanda causado por un crecimiento de la población sería, por ejemplo, uno de los objetivos que cumpliría el modelo. En la **Figura 47** se muestran las presiones mínimas y máximas alcanzadas en cada nudo a lo largo de todo el periodo simulado. En general, la variabilidad en media de las máximas parece ser un poco superior al de las mínimas. En el caso de estas últimas, se cree que dicha variabilidad es de una holgura manejable para la toma de decisiones, en tanto que su reducida acotación proporciona seguridad. Por otra parte, en la misma **Figura 48** puede observarse por separado la ubicación de aquellos puntos en los que, en alguna de las configuraciones del sistema encontradas, o la presión mínima está por debajo de 5 mca y/o la máxima por encima de 20 mca. La distribución de los mismos es acorde a lo esperado:

- Los nudos con presiones mínimas relativamente bajas están alejados de los puntos de producción o tienen una cota alta con respecto a los puntos de su alrededor. Dichas presiones mínimas seguramente se alcancen en horas del día con consumo alto.
- Por su parte, los nudos con presiones máximas relativamente altas se encuentran o cerca de los puntos de producción o bien en zonas de cota baja, dándose éstas presiones bajas en horas valle de consumo.



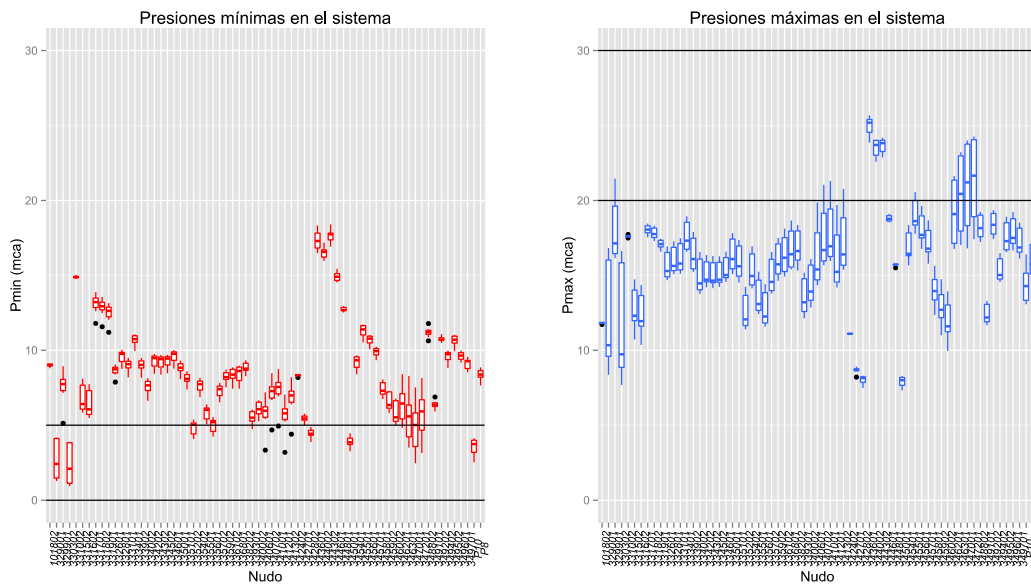


Figura 47: Resultados de presiones mínimas y máximas obtenidas con las distintas configuraciones encontradas del modelo.



Figura 48: Se encuentran resaltados, en la figura de la izquierda, los puntos en los que se dan presiones mínimas por debajo de 5 mca y, en la de la derecha, los de presión máxima por encima de 20 mca.

Para finalizar este apartado, se presentan en la **Figura 49** los gráficos con los resultados obtenidos en los puntos en los que se dispone de registro de presiones, pero que han sido descartados por distintas razones. Se recuerda que P1 se ubica alejado del punto inicial del ramal al que pertenece, siendo este último el que está incluido en el esquema simplificado de la red y al que se asocian los resultados mostrados. Por su lado, los registros de P7 y P9 están comprendidos en época otoño-invierno. La modulación mostrada en las series

observada y simulada es innegablemente parecida, en consecuencia, la conclusión que podría extraerse de dichos gráficos es que la bondad del ajuste del modelo resulta razonablemente buena, si bien salvando las distancias, en tanto que, por ejemplo, tendría más sentido que los valores de las series de P9 y P7 fueran mayores que los simulados. El resultado obtenido en P1 es el más satisfactorio, ya que los valores simulados siguen la misma modulación que los observados, pero son más altos que éstos, lo cual es coherente con el hecho de que se producen pérdidas de carga entre el punto inicial del ramal y el punto en el que se han realizado mediciones.

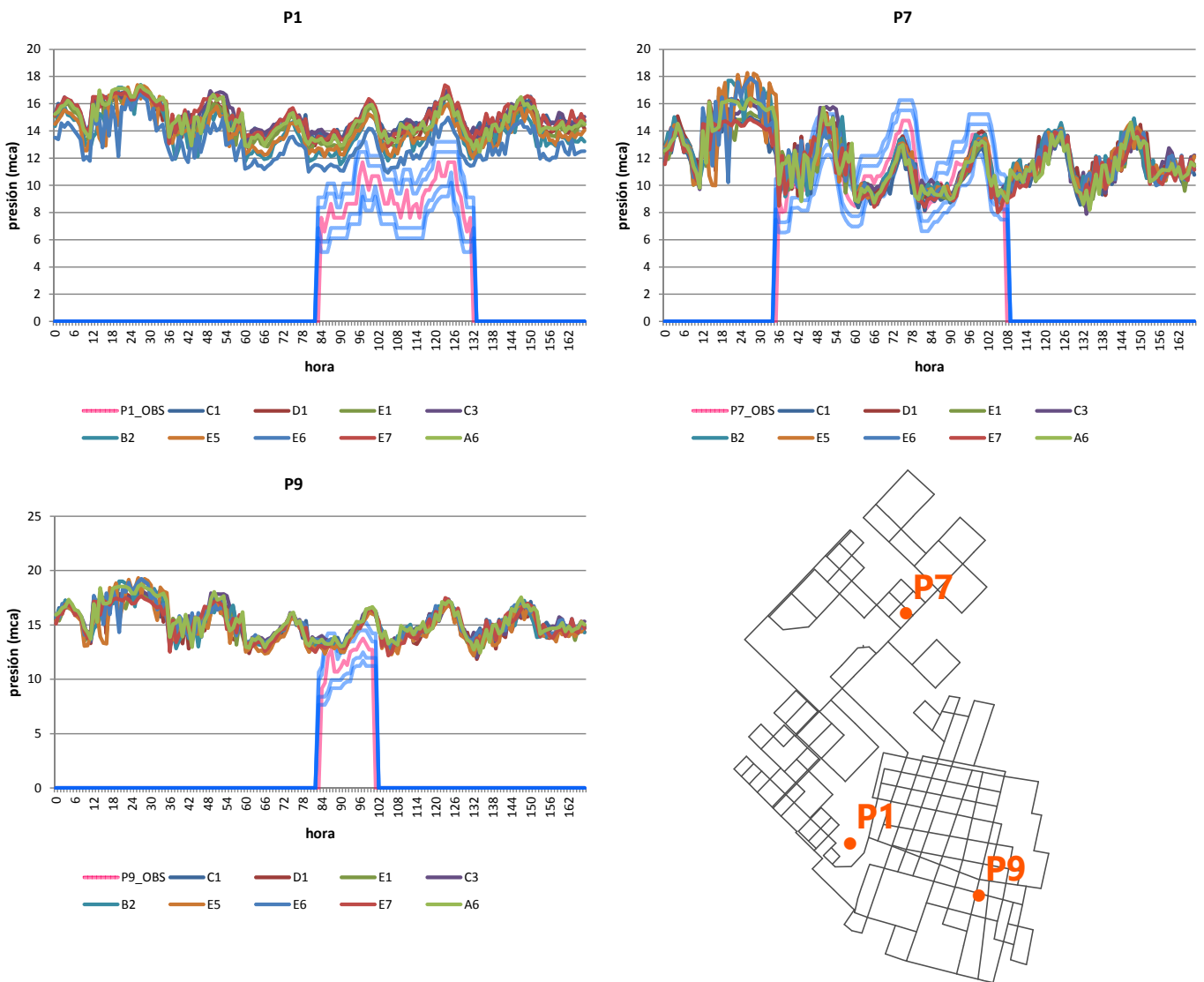


Figura 49: Gráficos de resultados correspondientes a los puntos que disponen de presiones observadas y figura en la que puede consultarse su ubicación, descartados por distintas razones.

4. Conclusiones y futuras líneas de investigación

Con respecto al caso de estudio, la primera y principal conclusión es que los resultados obtenidos no son fiables. Hay tres razones detrás de la misma:

- Los inputs de información para la calibración son insuficientes. Esto se afirma en base a las recomendaciones de algunos autores (Allen, 1987), que aconsejan realizar medidas de presión en un 15-20 % de los nudos del modelo, lo cual se traduce para una población de menos de 30000 habitantes en llevar a cabo tomas de datos sobre presión en torno a 30 puntos, distribuidos estratégicamente a lo largo de la red. Aparte, sería necesario disponer de algunas mediciones para comprobar la bondad del ajuste. Por otro lado, debería conocerse al menos el caudal inyectado a red en todos los puntos de producción.
- La metodología empleada y las hipótesis adoptadas para soslayar la limitación de datos no están validadas. Es decir, no existen suficientes experiencias ni tests de contrastación realizados a las mismas como para afirmar que los resultados obtenidos y las conclusiones derivadas son fiables.
- El propio trabajo de post-proceso de información realizado en el presente estudio no ha sido tampoco validado por un responsable de la red con conocimientos profundos sobre la misma. Que dicho trabajo se haya ejecutado en su totalidad por una persona aumenta la probabilidad de error.

Por tanto, queda en evidencia que la falta de información es uno de los primeros escollos que se deben superar. Conseguir mediciones del caudal inyectado por bombeo directo en el sector 3 y extender los registros de presión ya existentes a toda la semana para la época primavera-verano, serían primeros pasos en la buena dirección. También resultaría de gran interés obtener datos sobre los otros sectores, especialmente del 2, en relación a caudales inyectados y presiones observadas, de forma que podría llevarse a cabo su modelación. Además, cabría conocer con más precisión las conexiones entre tanques y pozos, así como los niveles dinámicos a los que trabajan estos últimos y la bomba instalada en el pozo 10, si se quiere intentar modelar con mayor grado de detalle los puntos de producción. Añadir para terminar, que en la medida de lo posible debería intentarse aumentarse la cantidad de puntos donde se toman registros de presión a lo largo de toda la red.

Por otro lado, todas las soluciones consideradas como admisibles coinciden en una cosa muy importante: el modelo, tal y como está elaborado, no puede reproducir a la vez la modulación que sigue la medición de caudal inyectado desde el tanque nro. 35 y la modulación de las presiones observadas el jueves y viernes a la media noche. Principalmente, esto se debe a que las series de caudal inyectado y presiones observadas

para la calibración resultan ser a primera vista incoherentes, dado que se produce un pico de caudal inyectado a la vez que todas las presiones observadas muestran un aumento progresivo (**Figura 40** y **Figura 41**, apartado 3.8, series observadas). Una posible explicación radica en que existe algún fenómeno particular de este sistema que no se ha modelado correctamente. Llegados a este punto, cabe señalar que pasadas las 22:00-23:00 se suele abrir la VRP ubicada en cabecera de red para dar presión al sistema y así intentar llenar la totalidad de los tanques domiciliarios. Efectivamente, en la **Figura 24** se pueden consultar los repuntes en la cota de energía aguas abajo de la VRP generados por el "bypass" de la misma entre las 23:00 y la 1:00. Si se tiene en cuenta la fuerte presencia de tanques domiciliarios en esta red, cobra sentido físico el ascenso coincidente de presión observada y caudal inyectado desde el tanque nro. 35. El modelo elaborado considera la demanda base de los nudos de consumo independiente de la presión y los indicios parecen indicar que esta aproximación es insuficiente:

- Los martes, día en que no existen registros de presión, la modulación del caudal se consigue aproximar de forma fehaciente (**Figura 40**) a costa de disminuir al máximo el caudal independiente de la presión e incrementar el caudal dependiente de la presión, que en el modelo elaborado se ha representado como una fuga a través de los emisores implementados en EPANET.
- Se han encontrado soluciones cuyo volumen fugado está por encima del 50% del volumen total inyectado. Dicho nivel de fugas no es improbable, pero queda la pregunta de si parte del mismo no será realmente consumo dependiente de la presión, gracias al cual se consigue emular el caudal inyectado desde el tanque nro. 35.

Debido a ello, se cree que es conveniente intentar modelar parte de la demanda como dependiente de la presión, para lo cual se hace necesario disponer de mayor cantidad de información sobre la distribución de tanques domiciliarios a lo largo de la planta urbana.

En relación al tratamiento dado al consumo, sería aconsejable verificar que aquellas parcelas que quedan dentro de la planta urbana consolidada y están clasificadas como consumos importantes, realmente lo son, ya que es probable que el procedimiento masivo aplicado a los datos sobre la entidad parcela en este sentido encierre errores. Además, se deberían focalizar esfuerzos en aclarar qué viviendas fuera de la planta urbana consolidada disponen de conexión domiciliaria a la red pública. Los apuntes realizados anteriormente están asociados a la metodología aplicada, no obstante, se ha de resaltar que la solución ideal al desconocimiento sobre el consumo es incorporar contadores en Nogoyá como una norma y no como una excepción, aunque ello conlleve abundante pedagogía política y social, ya que los beneficios en ahorro de agua (hasta un 30%) y, en consecuencia, de energía, son considerables (Vela, 1996).

Para finalizar las conclusiones concernientes al caso de estudio, destacar que no se debe olvidar que este modelo es una fotografía del sistema en un determinado momento. Así pues, la información contenida en el mismo tiene que ser actualizada constantemente para poder afirmar que es una representación de la realidad cuando vaya a ser empleado. Sería de interés que los avances realizados por almacenar la información sobre la red en capas SIG sean aprovechados y dichas actualizaciones se lleven a cabo sobre el mismo. En relación al proyecto iniciado en SIG, queda mucho trabajo por hacer: crear nuevas capas para cada elemento de la red, tratar ésta como tal a través de una aplicación SIG, elaborar una base de datos para asociar los partes de roturas a su respectiva tubería o disponer de un registro de acometidas por tubería, en el que se contabilice las viviendas con tanque domiciliario y se especifique el tipo de parcela según las clases establecidas, constituyen algunas de las iniciativas de utilidad.

A continuación, se detallan las conclusiones alcanzadas sobre la metodología:

1. En el problema planteado intervienen un gran número de variables, en consecuencia, el algoritmo genético se enfrenta a una cantidad de combinaciones inconmensurable. A esto se añade que no hay suficientes datos disponibles para la calibración, por tanto, existirán infinidad de posibles soluciones que emulen de forma admisible dichos datos. Algunas de estas soluciones no tendrán sentido, por lo que se cree necesario acotar el espacio de las mismas a través de las restricciones aplicadas a las variables. Si bien, como contrapartida se genera la posibilidad de descartar alguna solución factible. A colación de esto, se cree que establecer los límites de las variables es una decisión que debe ser tomada habiendo realizado una cantidad de pruebas considerable.
2. Las bandas de resultados, proporcionados por las distintas configuraciones del sistema, están sujetas también al número de simulaciones llevadas a cabo.
3. El hecho de que la *demand base* de los nudos de consumo no se situara dentro del intervalo más estricto, podría interpretarse como un indicio de que su correspondiente término en la función objetivo no ha tenido suficiente peso. Sería interesante llevar a cabo pruebas restringiendo los rangos de variación de dicha variable, acotándolos a valores más próximos a la media establecida en el apartado 3.5.3.1.2 para cada grupo de nudos de consumo.
4. El clústering de tuberías (rugosidad) y nudos (coeficientes de descarga) no ha tenido un reflejo en el modelo, ya que, por ejemplo, grupos con tuberías más antiguas y de menor diámetro no han obtenido necesariamente los valores más altos de rugosidad.

En línea con lo anteriormente expuesto, se creen pertinentes los siguientes estudios para complementar esta investigación:

- Simular la situación en cuanto a información limitada en una población en la que sí pueda elaborarse un modelo de la red aplicando los procedimientos

convencionales, de forma que los resultados obtenidos con la metodología presentada tendrían con qué compararse y sería posible evaluar el error que se comete empleando la misma. También cabría la posibilidad de validarla y de comprobar algunas de las hipótesis de partida asumidas como ciertas.

- Estimar la sensibilidad de cada una de las decisiones tomadas a lo largo del proceso de modelación, en un intento de responder preguntas como: ¿cuál es la repercusión del umbral elegido sobre el número mínimo de parcelas por nudo de consumo y la distribución adjudicada a éstos? ¿de qué manera afecta el número de clústeres seleccionado cada vez que se ha aplicado esta técnica?
- Llevar a cabo en distintos casos de estudio un mayor número de pruebas en la fase de calibración, para establecer directrices fundamentadas acerca de los límites de las variables y la cantidad total de simulaciones a ejecutar.

Seguidamente, se cierra este capítulo con la propuesta de tres líneas de investigación generales:

- Componer una aplicación en la que, dadas las distintas configuraciones del sistema encontradas, se puedan crear escenarios que generen modificaciones en todas ellas y, a continuación, para cada uno proporcione resultados análogos a los expuestos en el apartado 3.7.
- Encontrar nuevas metodologías que mediante la gestión y acotación de la incertidumbre permitan disminuir las necesidades de información.
- Desarrollar software SIG libre específico para servicios de abastecimiento de agua, así como herramientas análogas a Evolver.

Referencias

- Allen, R. (1987). *Network analysis- the real story*. IIIrd International Conference. Computer applications for water supply and distribution. Ed. Leicester Polytechnic, Reino Unido.
- Bartolín, H. (2013). *Confección de modelos de redes de distribución de agua desde un SIG y desarrollo de herramientas de apoyo a la toma de decisiones*. Tesis Doctoral por la Universitat Politècnica de València.
- Bosque J. 1992. *Sistemas de Información Geográfica*. Ediciones Rialph S.A., España.
- C.-C. Chang y C.-J. Lin (2001). *LIBSVM: A Library for Support Vector Machines*.
- DSAPSBPER, 2012. *Diagnóstico del Sistema de Agua Potable y Saneamiento Básico de la Provincia de Entre Ríos*, Consejo Federal de Inversiones.
- Echávez, G. 1997. *Increase in losses coefficient with age for small diameter pipes*. J. Hydraul. Eng.123:157-159.
- Idelchik, I. E. 1994. *Handbook of hydraulic resistance*, 3rd Ed., CRC Press, Inc., Boca Raton, Fla.
- Kaufman, L., y Rousseeuw, P.J. (1990). *Finding Groups in Data: An Introduction to Cluster Analysis*. Wiley, New York.
- López, G., Fuertes, V., Ayza, M. (1996). *Modelización Matemática de una Red en Funcionamiento*. Ingeniería hidráulica aplicada a los Sistemas de Distribución de Agua (Ed. Cabrera et. al.), pp. 729-760, vol. II, España.
- Murtagh, F. y Legendre, P. (2014). *Ward's Hierarchical Agglomerative Clustering Method: Which Algorithms Implement Ward's Criterion?* Journal of Classification 31, pp. 274-295.
- Nash, J. E. and J. V. Sutcliffe. 1970, *River flow forecasting through conceptual models part I - A discussion of principles*. Journal of Hydrology, 10 (3), 282-290.
- Rossman, L. (2001). *Manual de usuario de Epanet 2*. Traducción al español de Fernández Martínez Alzamora, Grupo IDMH. Dep. Ingeniería Hidráulica y M.A., Universidad Politécnica de Valencia (UPV).
- SIGMUR, 2005. *Temario de las asignaturas Sistemas de Información Geográfica de la licenciatura de Ciencias Ambientales e Introducción a los Sistemas de Información Geográfica de la licenciatura de Geografía*, curso 2005-2006.

Takahashi, S., J. G. Saldarriaga, J.G., Vega, M.C. y Hernández, F. (2010). *Water distribution system model calibration under uncertainty environments*. Water Science & Technology, 10.1, pp. 31-38.

Tomás, J.; Valenti, R., Duarte, O., Graizaro, S. y Sione, H. 1999. *Aptitud del Agua subterránea con destino a riego en sectores de los departamentos Paraná, Diamante y Nogoyá de la provincia de Entre Ríos*. Serie Correlación Geológica 13, 279-286.

Vela, A., Andrés, M., López, P.A. (1996). *Auditoría y Mejora del Rendimiento de un Abastecimiento*. Ingeniería hidráulica aplicada a los Sistemas de Distribución de Agua (Ed. Cabrera et. al.), pp. 997-1048, vol. II, España.

Valero, S. (2010). *Arquitectura de Búsqueda Basada en Técnicas Soft Computing para la Resolución de Problemas Combinatorios en Diferentes Dominios de Aplicación*. Tesis Doctoral por la Universitat Politècnica de València.

Winarni, W. 2009. *Infrastructure Leakage Index (ILI) as Water Losses Indicator*. Technical Note. Civil Engineering Dimension, Vol. 11, No. 2, 126-134. ISSN 1410-9530 print / ISSN 1979-570X online.

ZAHER, 2008. *Estudios socioeconómicos de la sustentabilidad de los sistemas de producción y recursos naturales, n° 6. Zonas Agroecológicas Homogéneas, Entre Ríos*. Instituto Nacional de Tecnología Agropecuaria (INTA).

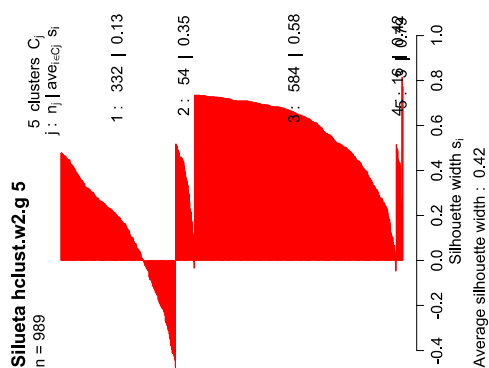
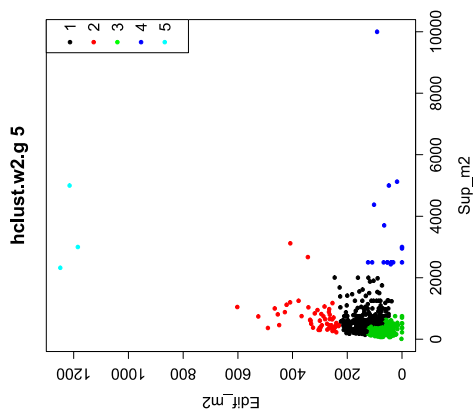
Mirkin, B. (2005). *Clustering for Data Mining: A Data Recovery Approach*. Taylor & Francis Group.

ANEJOS

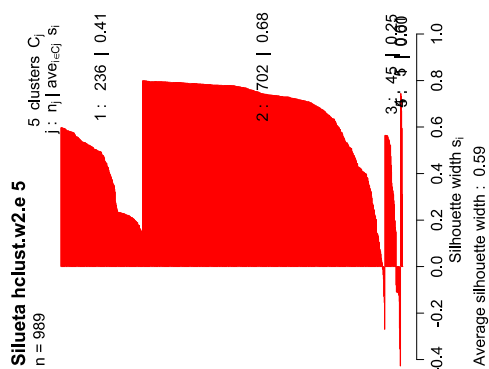
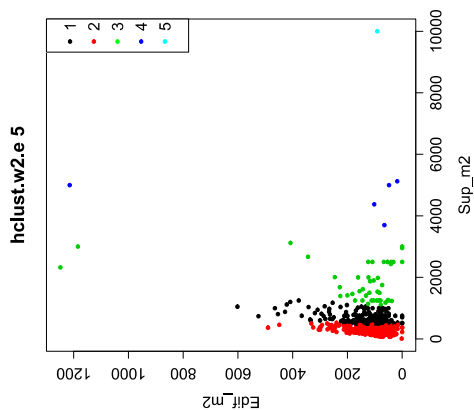
Anejo 1: Clústering de María Grande

Clústering jerárquico aglomerativo. Método ward.D2.

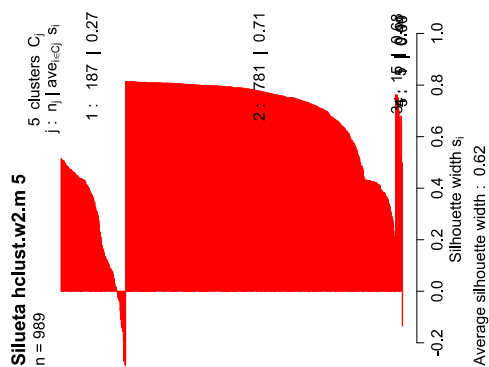
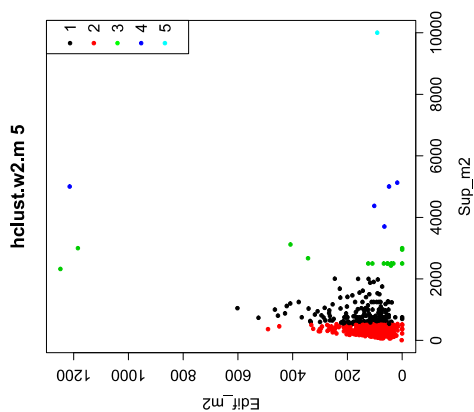
.Métrica Gower



Métrica Euclidiana

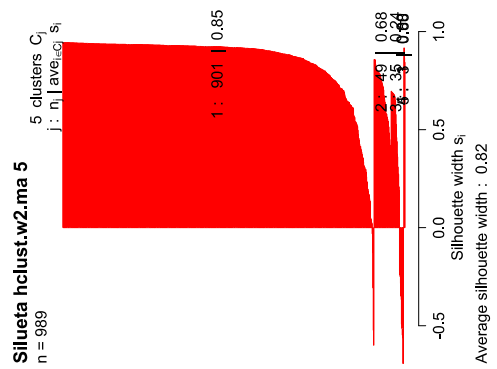
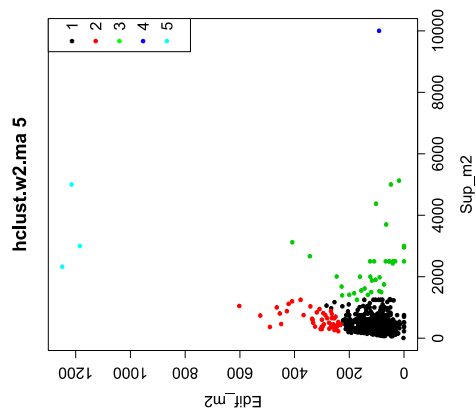


Métrica Manhattan

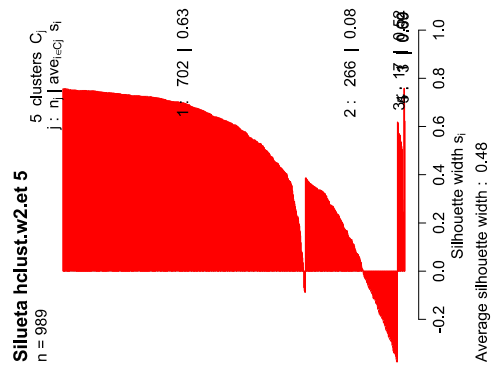
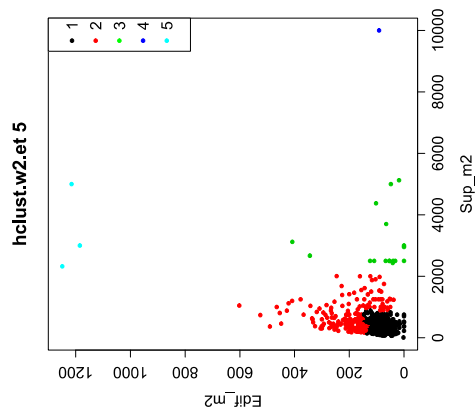


Clustering jerárquico aglomerativo. Método ward.D2.

Métrica Mahalanobis



Variables estandarizadas. Métrica Euclidiana.



Variables estandarizadas. Métrica Manhattan.

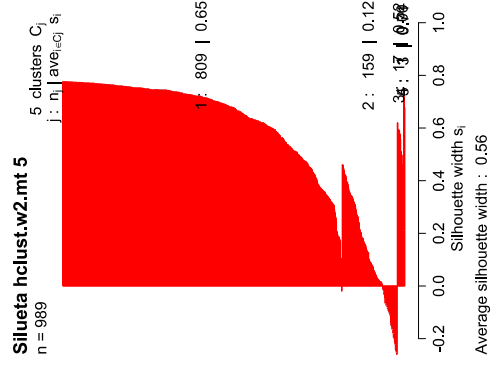
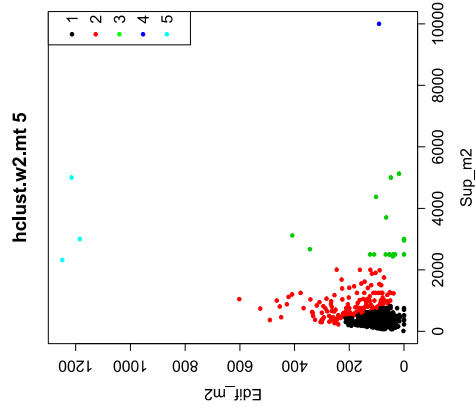
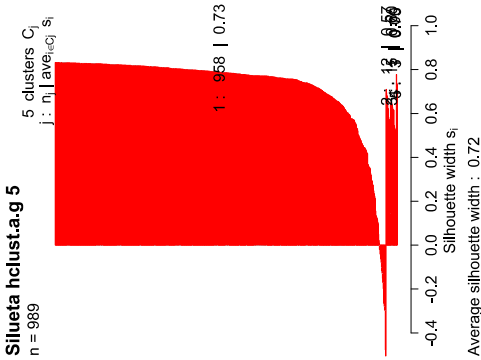
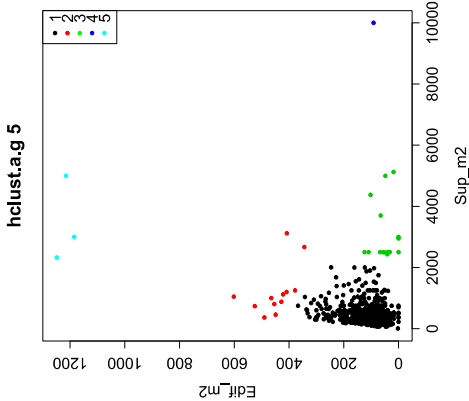


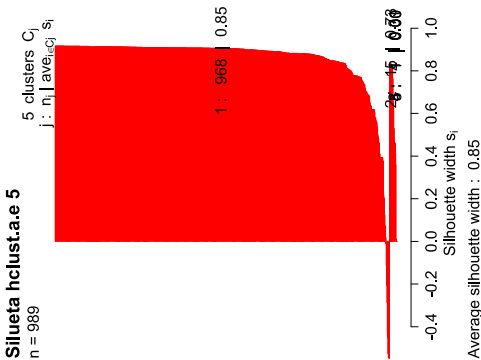
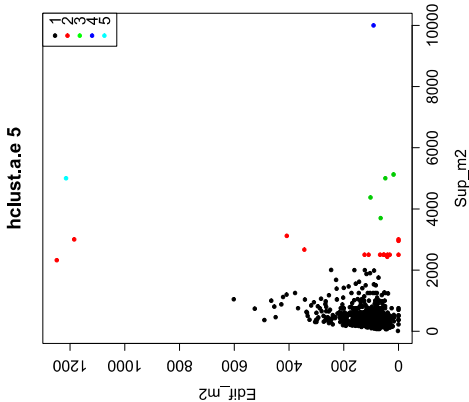
Tabla A1: Gráficos de dispersión Edif_m2 vs. Sup_m2 y de la silueta correspondiente al clustering jerárquico aglomerativo, método ward.D2, para 5 clústeres y cada métrica probada.

Clustering jerárquico aglomerativo. Método UPGMA.

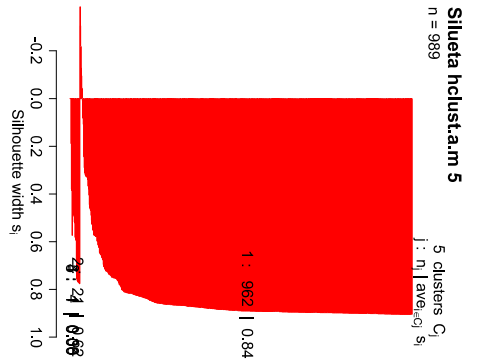
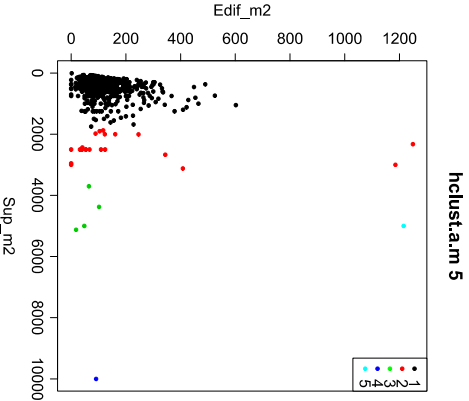
Métrica Gower



Métrica Euclidiana

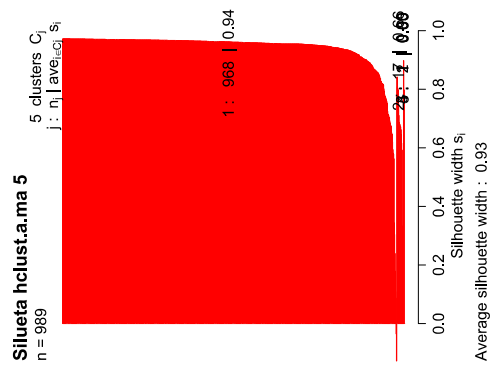
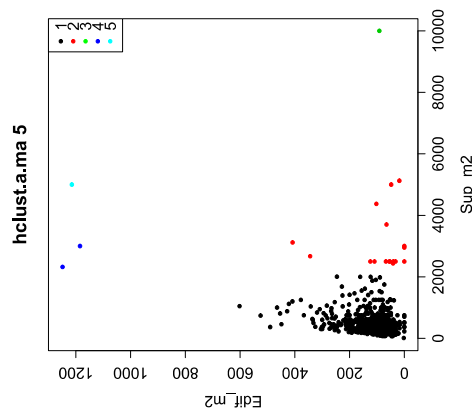


Métrica Manhattan

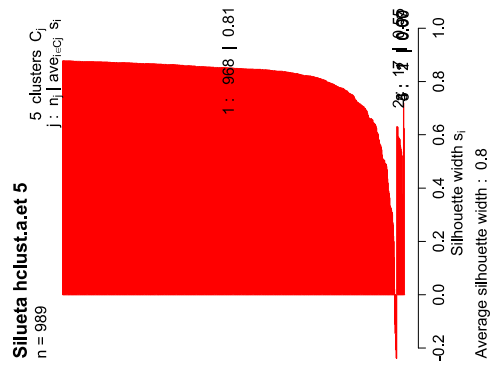
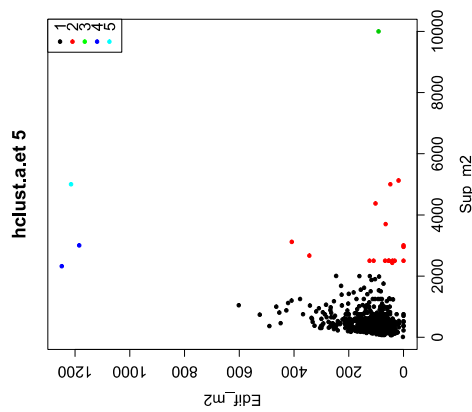


Clustering jerárquico aglomerativo. Método UPGMA.

Métrica Mahalanobis



Variables estandarizadas. Métrica Euclidiana.



Variables estandarizadas. Métrica Manhattan.

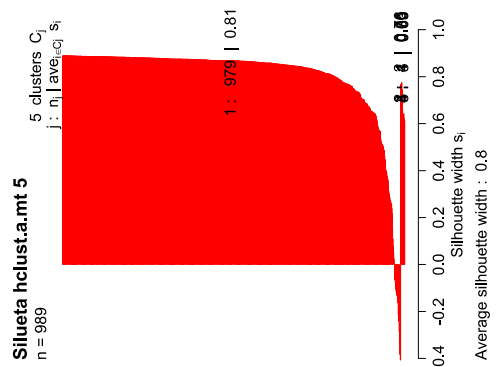
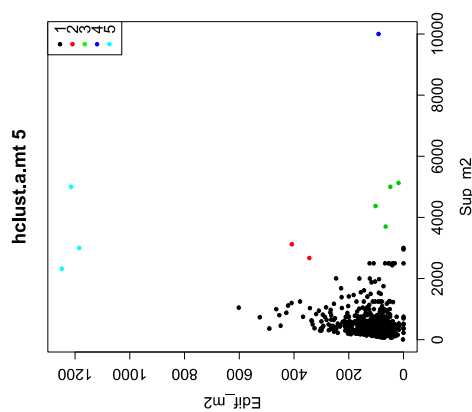
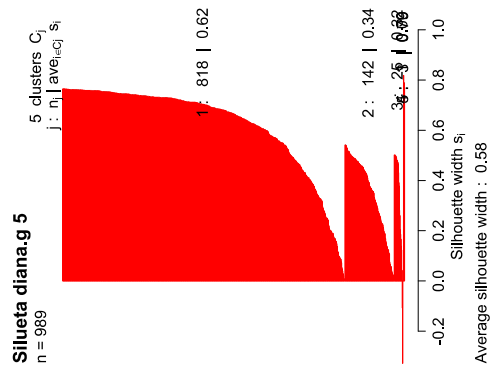
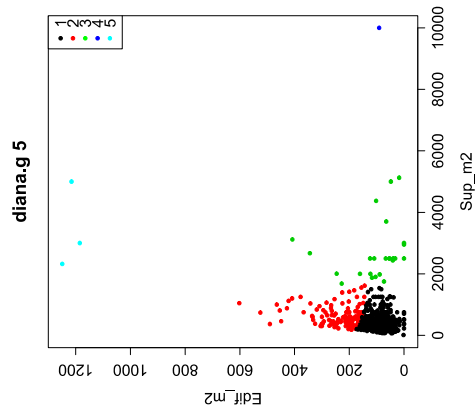


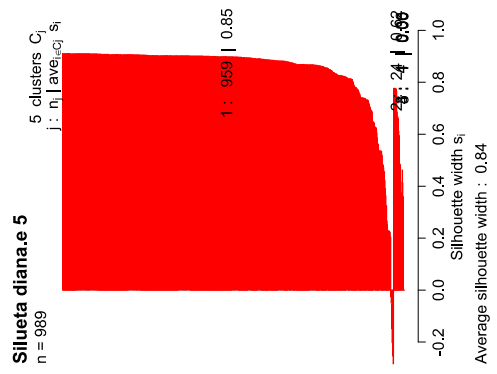
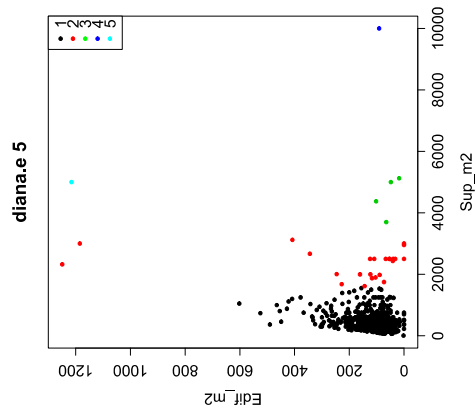
Tabla A2: Gráficos de dispersión Edif_m2 vs. Sup_m2 y de la silueta correspondiente al clustering jerárquico aglomerativo, método UPGMA, para 5 clústeres y cada métrica probada.

Clustering jerárquico divisivo DIANA

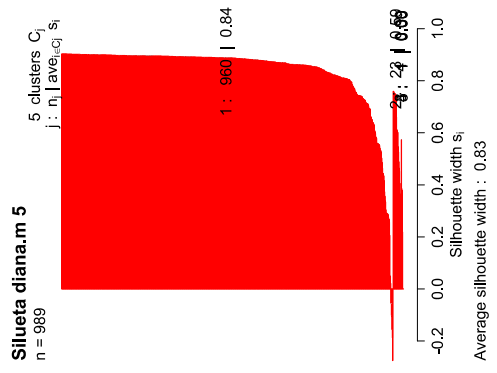
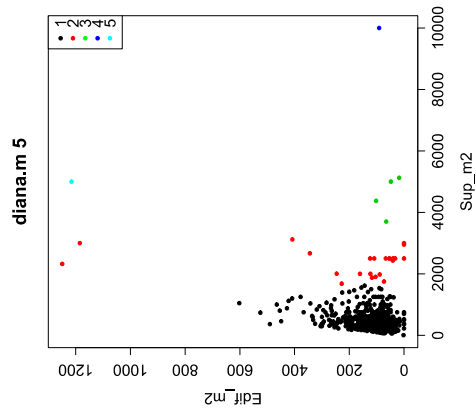
.Métrica Gower



Métrica Euclidiana

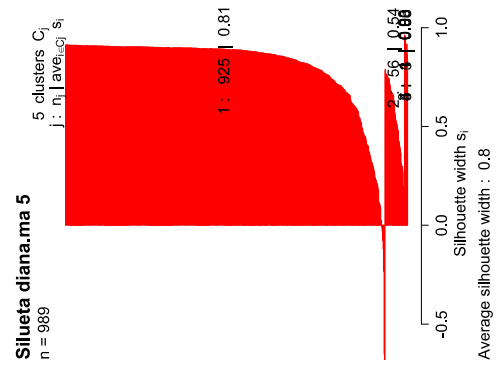
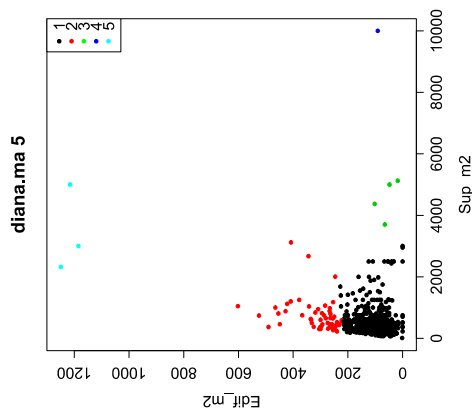


Métrica Manhattan

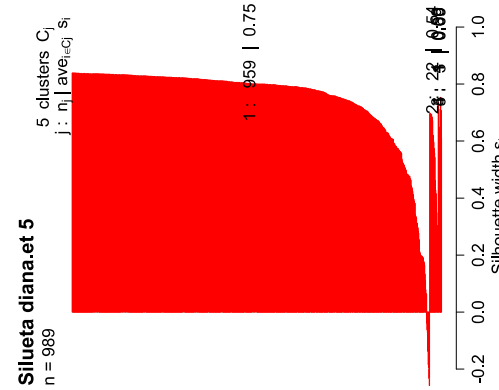
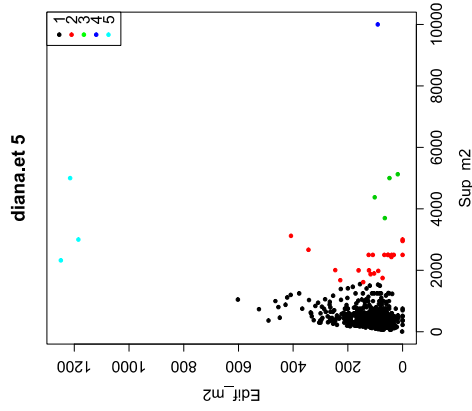


Clustering jerárquico divisivo DIANA

Métrica Mahalanobis



Variables estandarizadas. Métrica Euclidiana



Variables estandarizadas. Métrica Manhattan

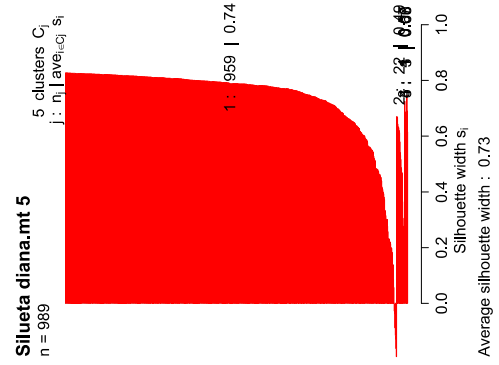
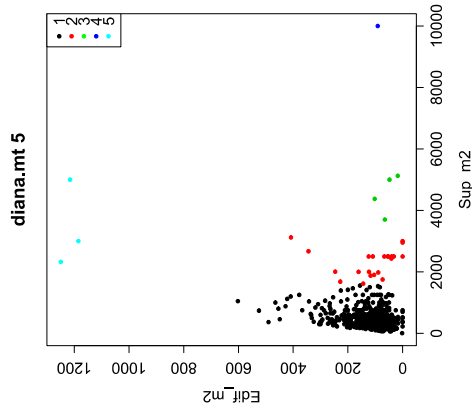
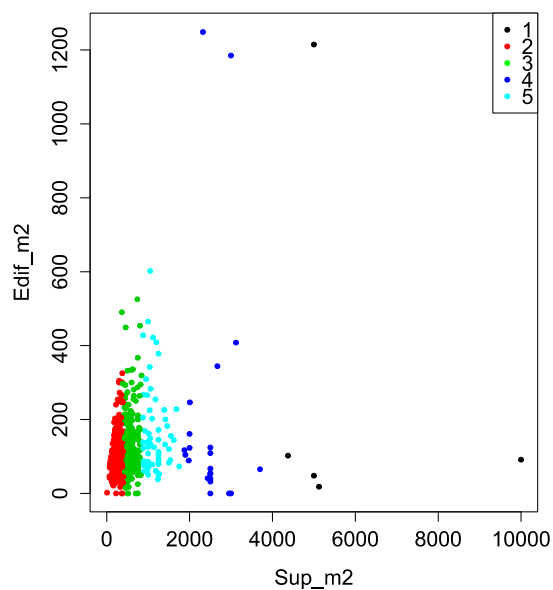


Tabla A3: Gráficos de dispersión Edif_m2 vs. Sup_m2 y de la silueta correspondiente al clustering jerárquico divisivo DIANA, para 5 clústeres y cada métrica probada.

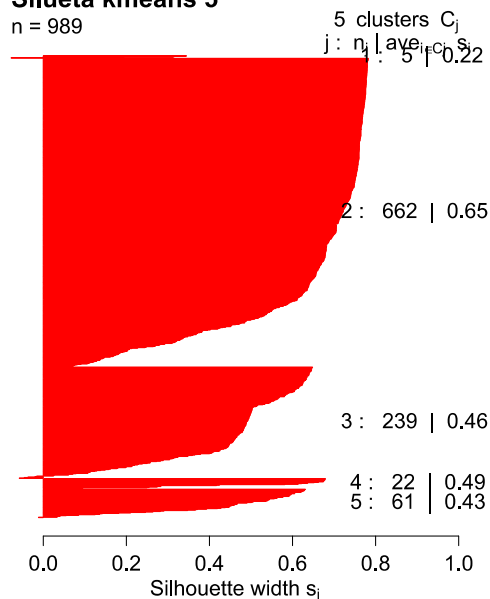
Clustering de partición KMEANS
Métrica Euclidiana

kmeans 5



Silueta kmeans 5

n = 989

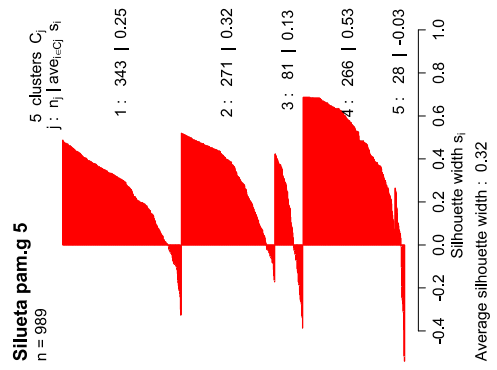
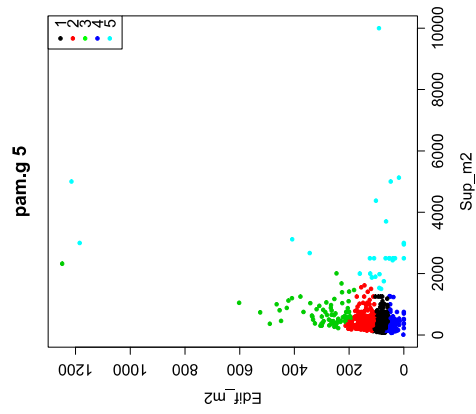


Average silhouette width : 0.58

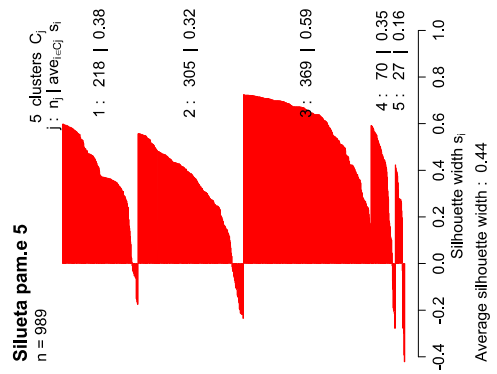
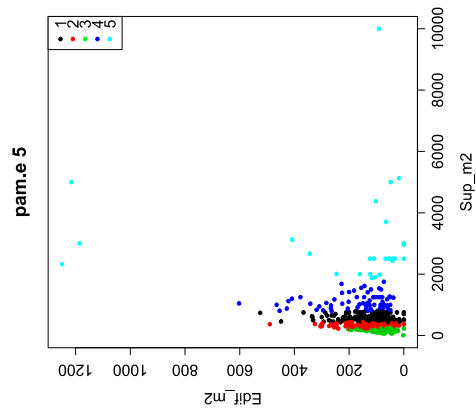
Tabla A4: Gráfico de dispersión Edif_m2 vs. Sup_m2 y de la silueta correspondiente al clustering de partición KMEANS, para 5 clústeres.

Clustering de partición PAM

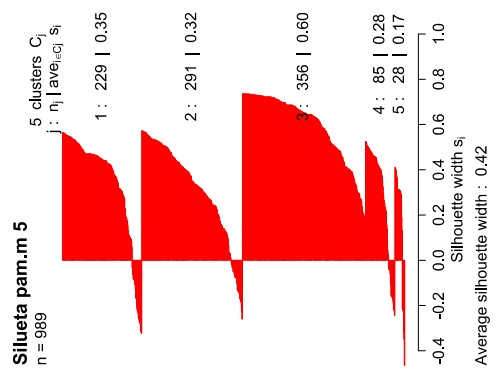
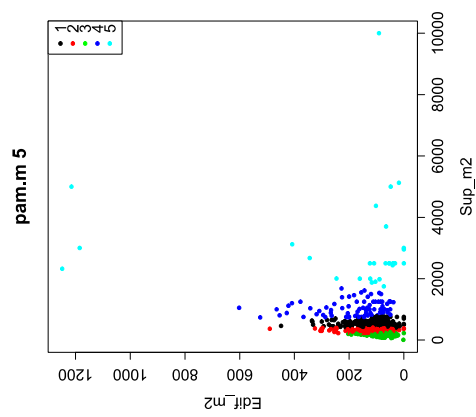
Métrica Gower



Métrica Euclidiana



Métrica Manhattan



Clustering de partición PAM

Métrica Mahalanobis

Variables estandarizadas. Métrica Euclidiana.

Variables estandarizadas. Métrica Manhattan.

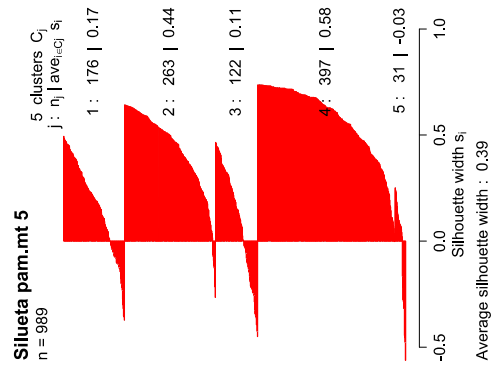
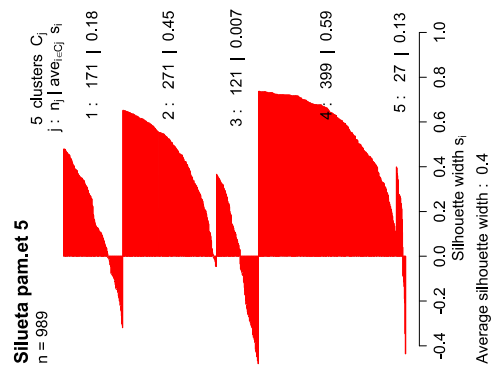
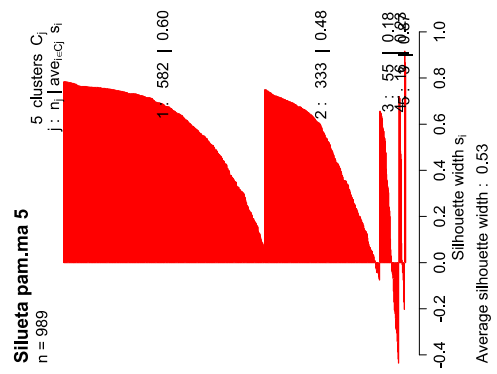
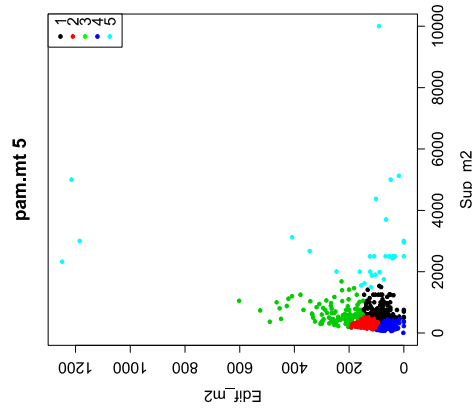
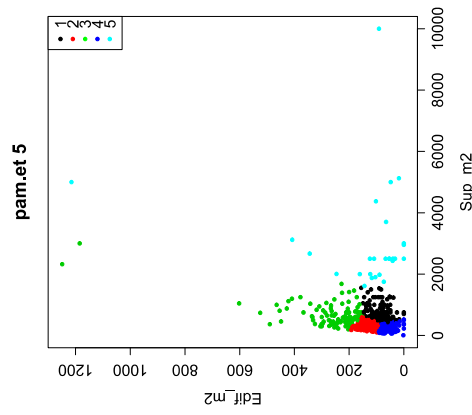
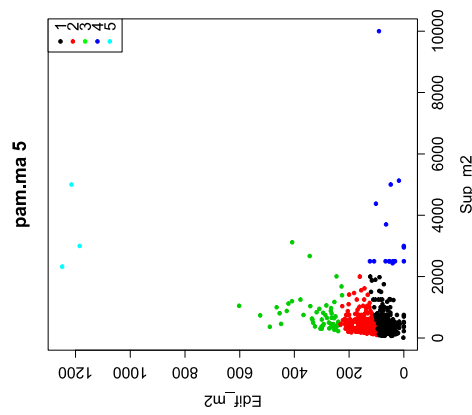


Tabla A5: Gráficos de dispersión Edif_m2 vs. Sup_m2 y de la silueta correspondiente al clustering de partición PAM, para 5 clústeres y cada métrica probada.

Anejo 2: Registros de presiones observadas distribuidas por la red de Nogoyá

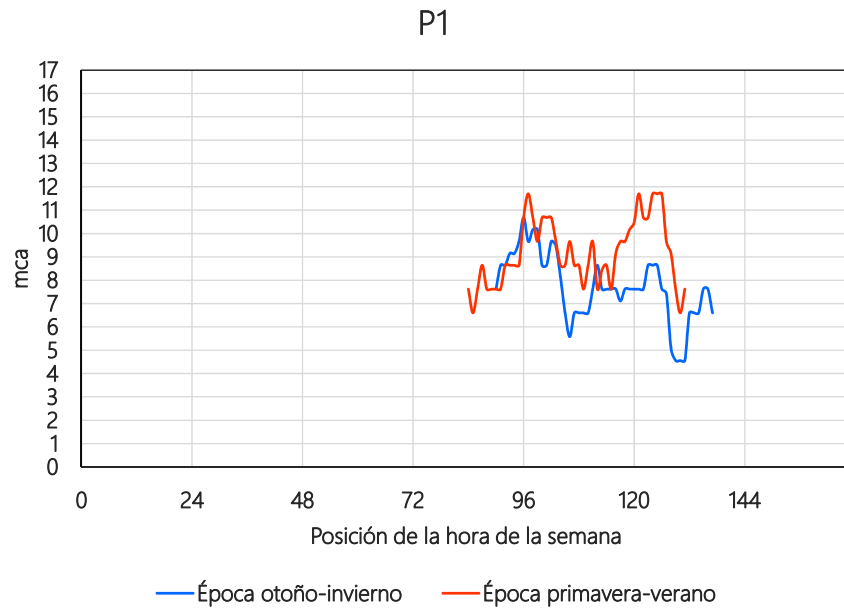


Figura A1: Registro de presiones para la época primavera-verano asociado a la medición P1, ubicada en el domicilio con dirección 17 de Agosto 1967. El medidor se instaló en una acometida conectada directamente a una tubería de distribución de diámetro 63 mm.



Figura A2: Registro de presiones para la época primavera-verano asociado a la medición P2, ubicada en el domicilio con dirección Avda. Marchini y Bv. Alberdi. El medidor se instaló en una acometida conectada directamente a una tubería de distribución de diámetro 160 mm.

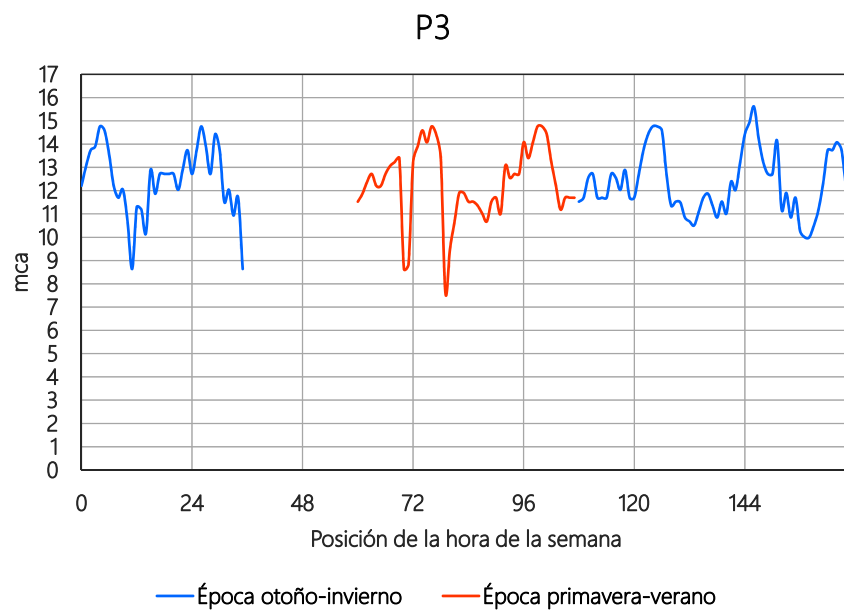


Figura A3: Registro de presiones para la época primavera-verano asociado a la medición P3, ubicada en el domicilio con dirección Bv. España 1346. El medidor se instaló en una acometida conectada directamente a una tubería de distribución de diámetro 125 mm.

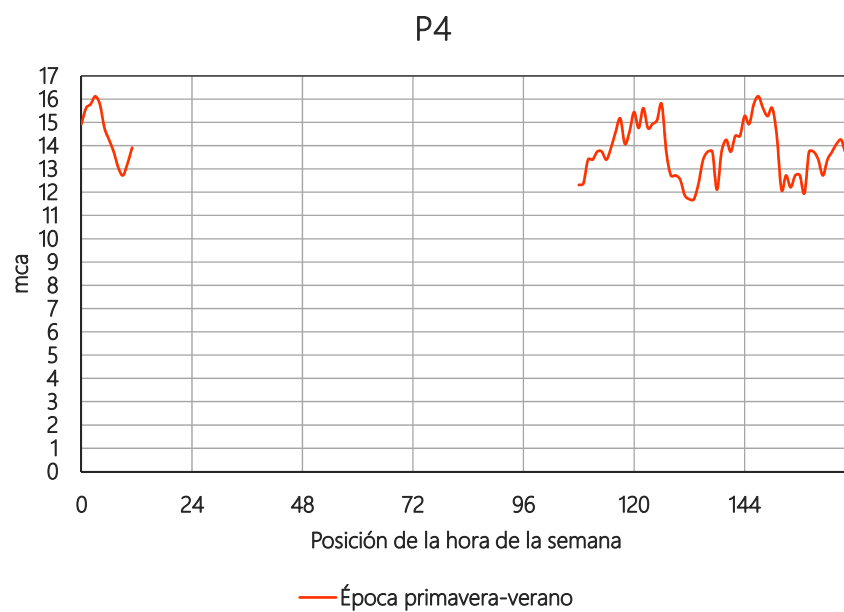


Figura A4: Registro de presiones para la época primavera-verano asociado a la medición P4, ubicada en el domicilio con dirección Bv. España y 25 de Mayo. El medidor se instaló en una acometida conectada directamente a una tubería de distribución de diámetro 200 mm.

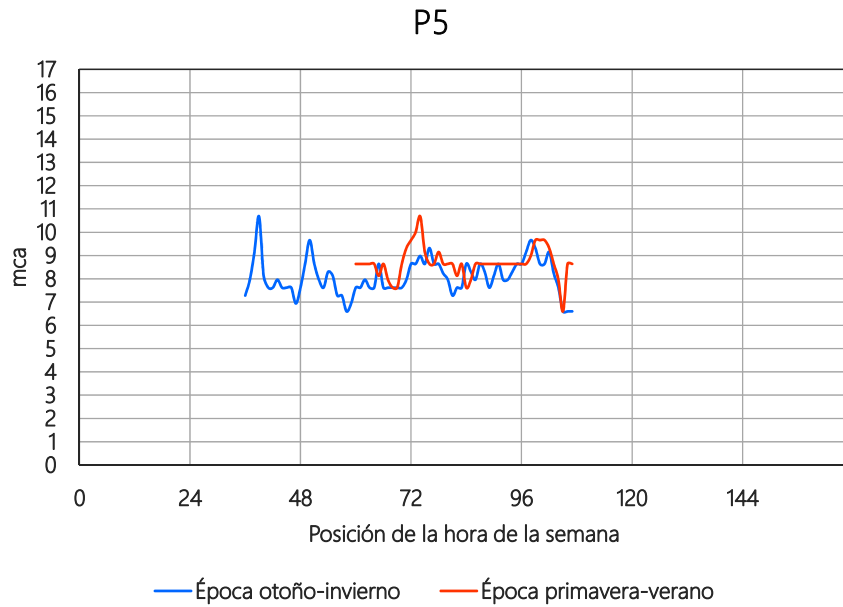


Figura A5: Registro de presiones para la época primavera-verano asociado a la medición P5, ubicada en el domicilio con dirección F. Gerald y Q. y Taboada 876. El medidor se instaló en una acometida conectada directamente a una tubería de distribución de diámetro 160 mm.

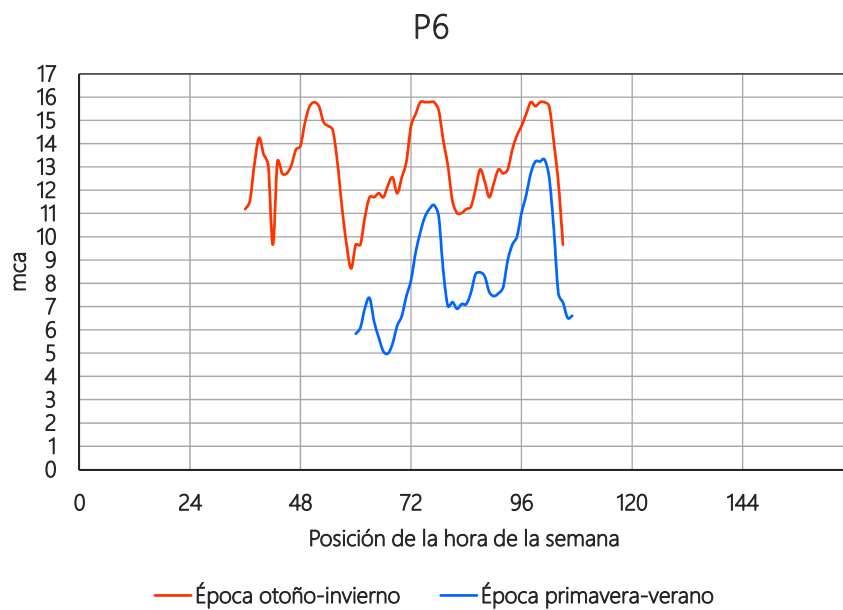


Figura A6: Registro de presiones para la época primavera-verano asociado a la medición P6, ubicada en el domicilio con dirección J.J. Paso y Marchini. El medidor se instaló en una acometida conectada directamente a una tubería de distribución de diámetro 63 mm.



Figura A7: Registro de presiones para la época primavera-verano asociado a la medición P7, ubicada en el domicilio con dirección Avda. Marchini y Colon. El medidor se instaló en una acometida conectada directamente a una tubería de distribución de diámetro 200 mm.



Figura A8: Registro de presiones para la época primavera-verano asociado a la medición P8, ubicada en el domicilio con dirección Avda. Marchini 1582. El medidor se instaló en una acometida conectada directamente a una tubería de distribución de diámetro 200 mm.



Figura A9: Registro de presiones para la época primavera-verano asociado a la medición P9, ubicada en el domicilio con dirección Q. y Taboada 876. El medidor se instaló en una acometida conectada directamente a una tubería de distribución de diámetro 200 mm.

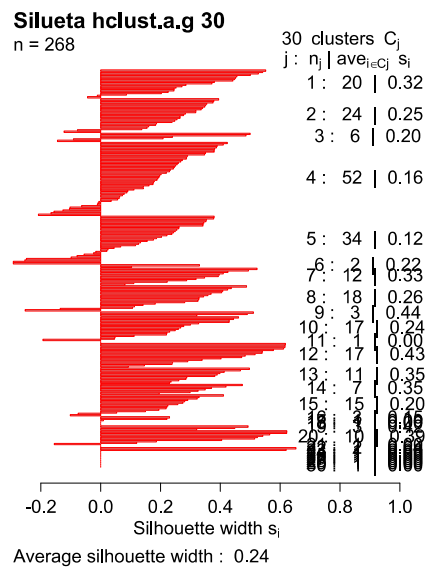
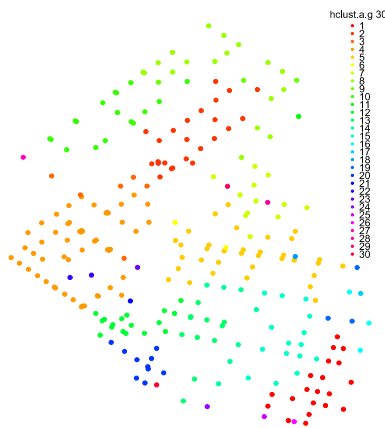
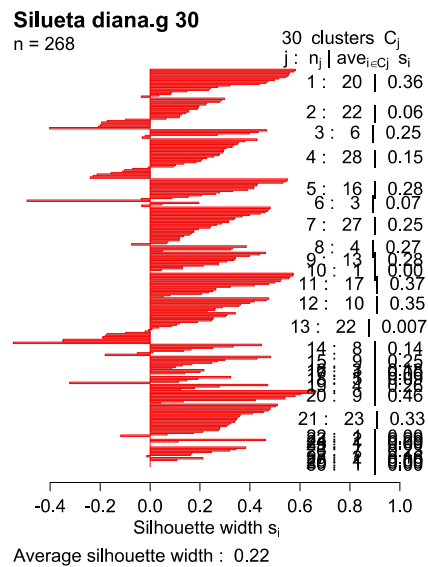
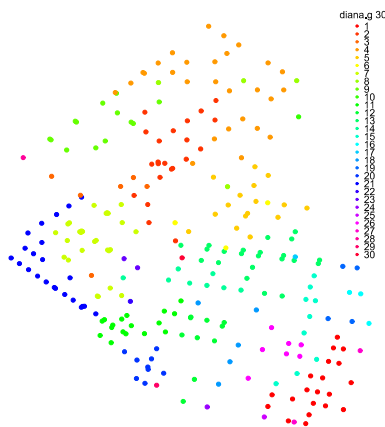
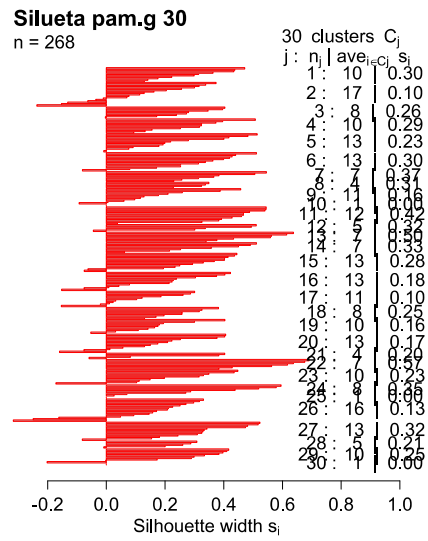
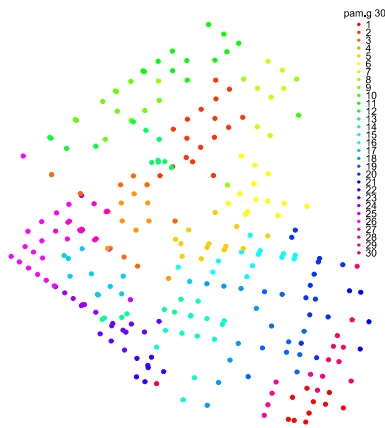


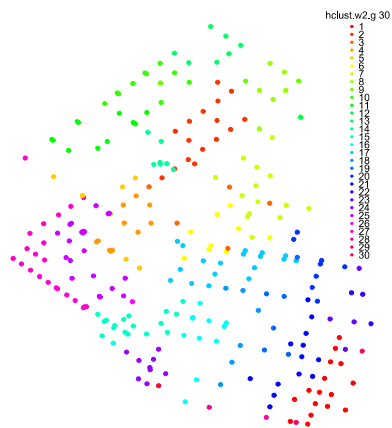
Figura A10: Registro de presiones para la época primavera-verano asociado a la medición P10, ubicada en el domicilio con dirección Ruta 12 y Victoria. El medidor se instaló en una acometida conectada directamente a una tubería de distribución de diámetro 63 mm.



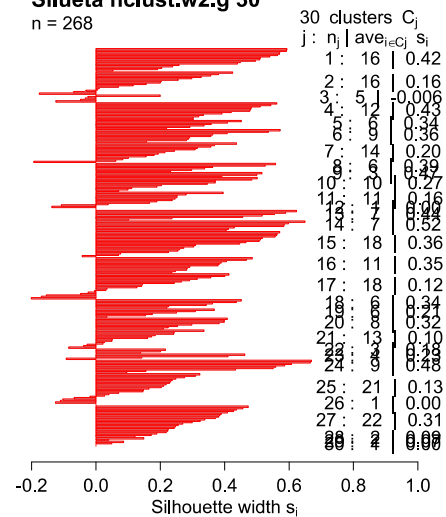
Figura A11: Registro de presiones para la época primavera-verano asociado a la medición P11, ubicada en el domicilio con dirección Yrigoyene/Parana y S. Oriz. El medidor se instaló en una acometida conectada directamente a una tubería de distribución de diámetro 175 mm.

Anejo 3: Clústering de tuberías y nudos

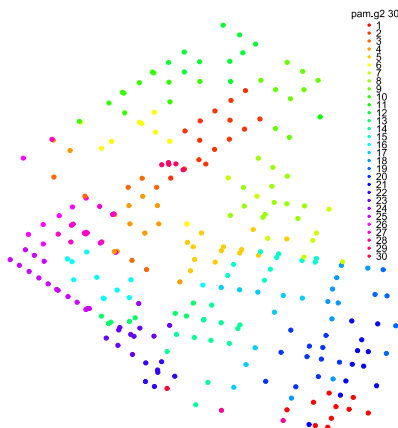




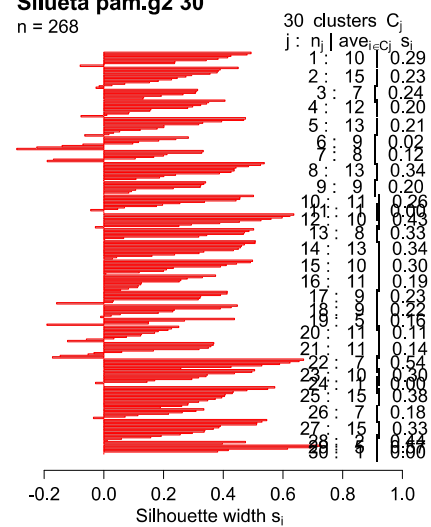
Silueta hclust.w2.g 30
n = 268



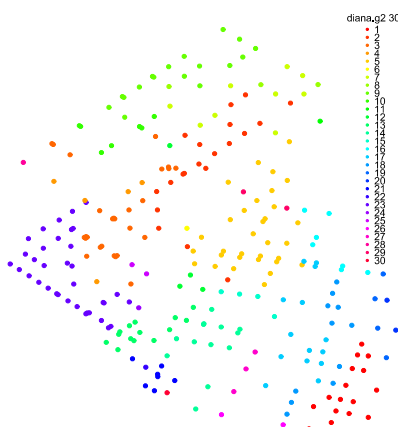
Average silhouette width : 0.27



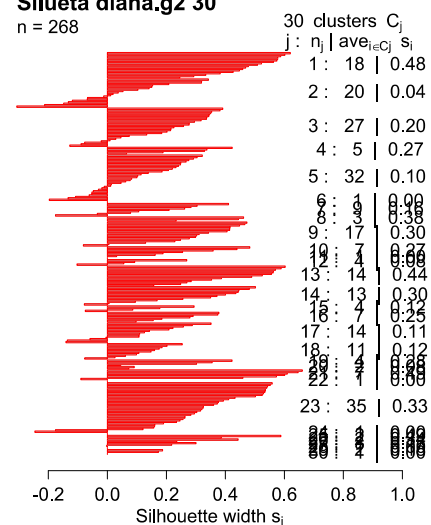
Silueta pam.g2 30
n = 268



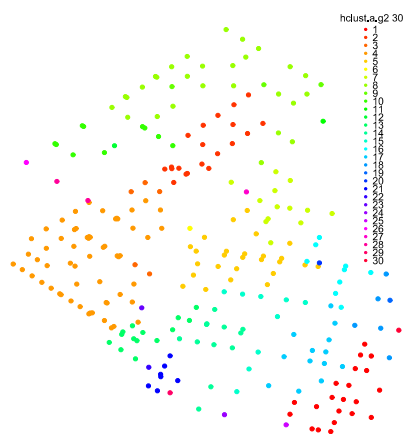
Average silhouette width : 0.26



Silueta diana.g2 30
n = 268



Average silhouette width : 0.24



Siluetas hclust.a.g2 30

n = 268

30 clusters C_j

j : n_j | $\text{ave}_{i \in C_j} s_i$

1 : 21 | 0.46

2 : 20 | 0.30

3 : 5 | 0.27

4 : 58 | 0.20

5 : 26 | 0.21

6 : 1 | 0.00

7 : 13 | 0.27

8 : 28 | 0.21

9 : 3 | 0.39

10 : 10 | 0.00

11 : 10 | 0.00

12 : 14 | 0.42

13 : 12 | 0.35

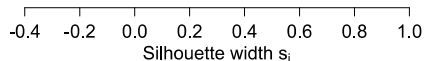
14 : 12 | 0.28

15 : 9 | 0.35

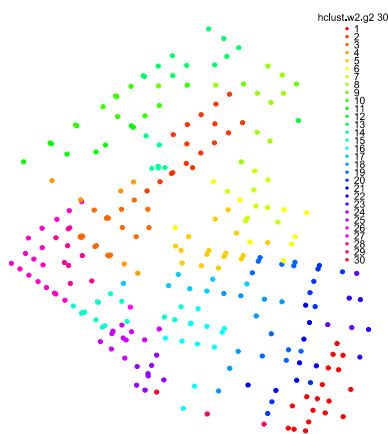
16 : 13 | 0.13

17 : 5 | 0.00

18 : 4 | 0.00



Average silhouette width : 0.26



Siluetas hclust.w2.g2 30

n = 268

30 clusters C_j

j : n_j | $\text{ave}_{i \in C_j} s_i$

1 : 18 | 0.46

2 : 17 | 0.15

3 : 19 | 0.32

4 : 6 | 0.23

5 : 16 | 0.29

6 : 9 | 0.02

7 : 8 | 0.38

8 : 10 | 0.30

9 : 10 | 0.10

10 : 11 | 0.09

11 : 6 | 0.40

12 : 15 | 0.14

13 : 15 | 0.14

14 : 12 | 0.38

15 : 7 | 0.16

16 : 8 | 0.17

17 : 11 | 0.20

18 : 8 | 0.24

19 : 11 | 0.17

20 : 10 | 0.46

21 : 10 | 0.17

22 : 11 | 0.42

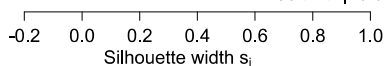
23 : 11 | 0.00

24 : 17 | 0.36

25 : 9 | 0.00

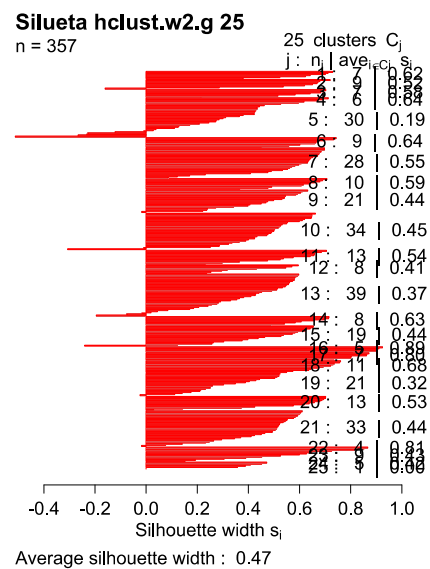
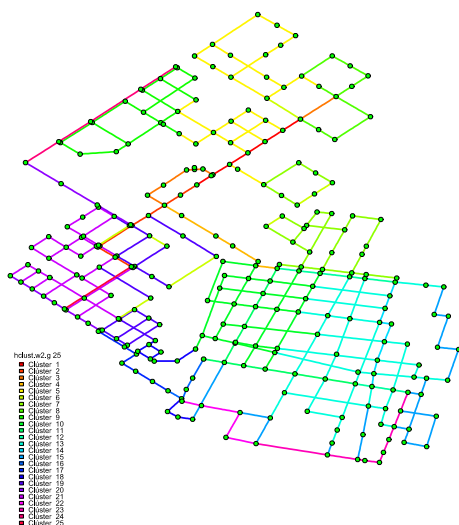
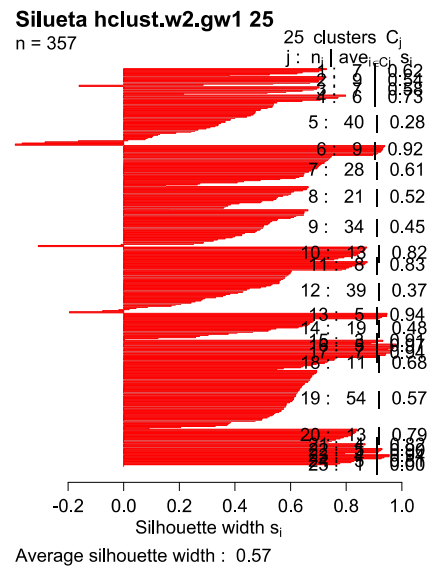
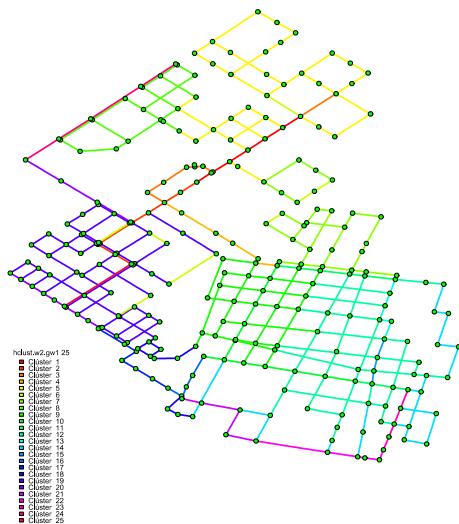
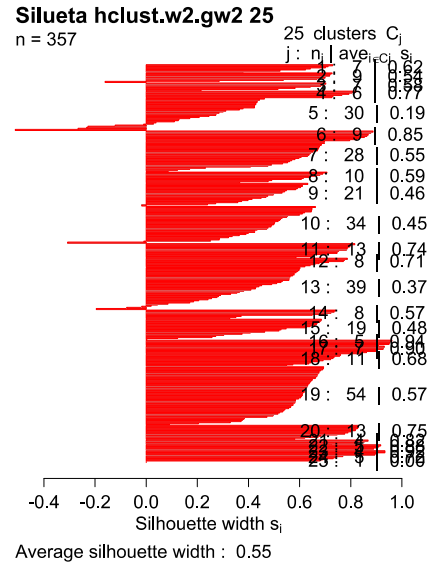
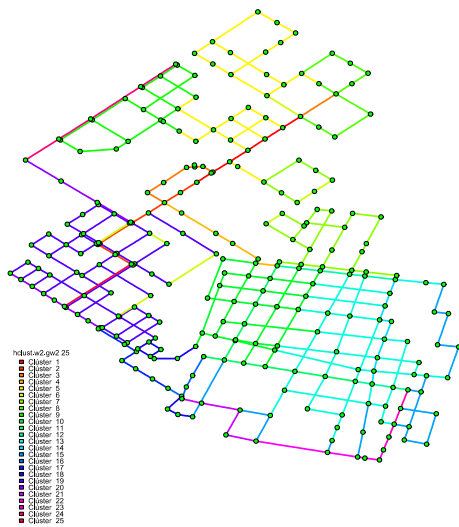
26 : 9 | 0.26

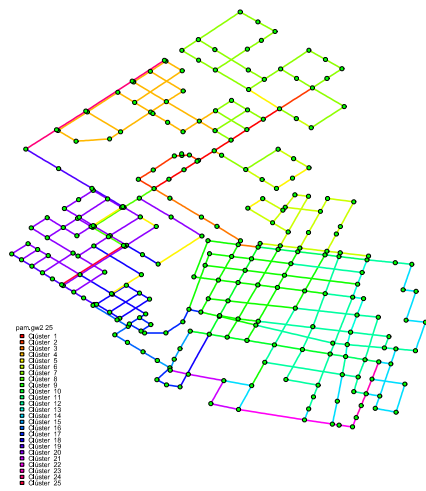
27 : 9 | 0.00



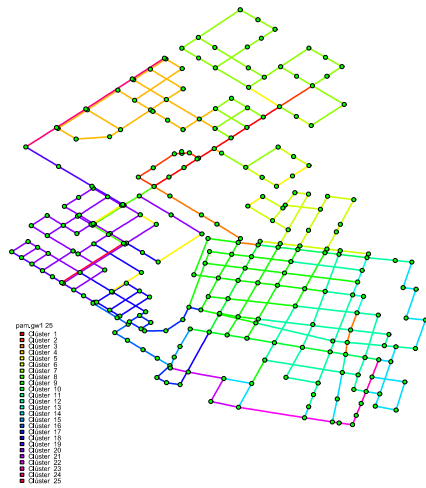
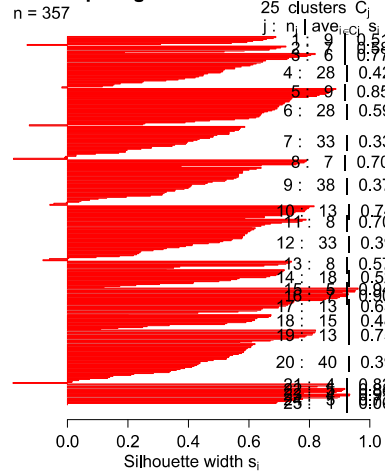
Average silhouette width : 0.28

Figura A12: Siluetas y gráficos de puntos asociados al clústering de nudos, a través de los cuales se pueden identificar los distintos grupos conformados según tipo de clústering y combinaciones de pesos (aplicados a las variables de entrada) para el número máximo de grupos considerado, 30.

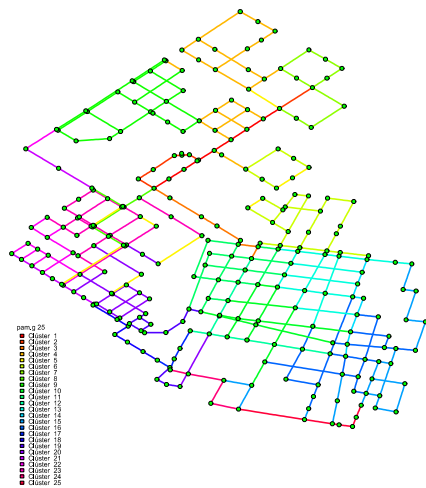
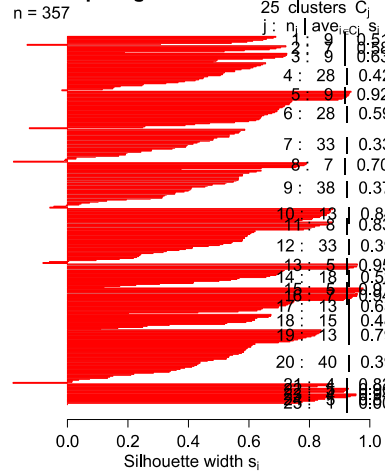




Silueta pam.gw2 25



Silueta pam.gw1 25



Silueta pam.g 25

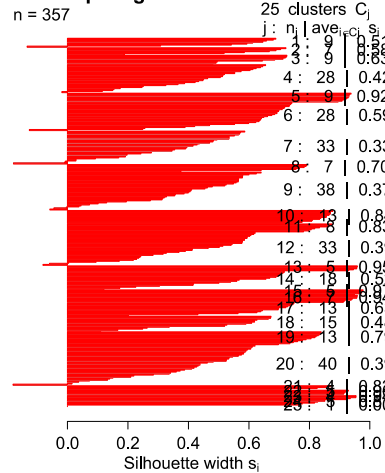


Figura A13: Siluetas y grafo de la red asociados al clústering de tuberías, a través de los cuales se pueden identificar los distintos grupos conformados según tipo de clústering y combinaciones de pesos (aplicados a las variables de entrada) para el número máximo de grupos considerado, 25.

Anejo 4: Calibraciones realizadas en Nogoyá

Prueba	Semilla	Tamaño población	Tasa de cruce	Tasa de mutación	Lím. Máx. Rugosidad	Lím. Máx. Coef. descarga	FO	P	Q	DBp	DBc	Criteri para
A1	162592248	125	0.4	0.15	0.35	0.2	4.67	0.73	3.49	0.36	0.09	P
B1	301889873	75	0.55	0.15	0.35	0.2	46.02	19.06	26.59	0.28	0.09	P
C1	958881702	50	0.6	0.1	0.35	0.2	2.07	0.49	1.33	0.16	0.09	P
D1	192410576	100	0.4	0.05	0.35	0.2	4.48	2.39	1.66	0.33	0.1	P
E1	100012295	150	0.6	0.1	0.35	0.2	2.67	0.66	1.55	0.24	0.22	NM
A2	188025998	50	0.6	0.15	0.25	0.1	111.52	76.35	34.71	0.29	0.17	P
B2	330068107	75	0.4	0.05	0.25	0.1	7.06	4.98	1.62	0.22	0.23	NM
C2	970176206	150	0.5	0.1	0.25	0.1	85.46	37.73	47.11	0.43	0.19	NM
D2	286838795	50	0.45	0.1	0.25	0.1	20.8	6.33	10.05	0.26	0.16	P
E2	117366498	100	0.6	0.15	0.25	0.1	71.54	41.74	29.38	0.29	0.12	P
D3	325435841	50	0.6	0.05	0.35	0.1	5.49	1.73	3.31	0.29	0.16	P
B3	379664826	75	0.45	0.15	0.35	0.1	34.87	11.7	22.7	0.23	0.24	P
E5	45590216	125	0.4	0.05	0.35	0.1	2.79	0.7	1.55	0.25	0.29	P
A5	250108873	125	0.45	0.15	0.35	0.1	68.3	44.67	23.08	0.27	0.28	NM
E6	62258466	125	0.6	0.05	0.35	0.1	2.44	0.49	1.5	0.27	0.17	P
E3	644482	75	0.4	0.15	0.25	0.2	20.47	3.03	17	0.27	0.18	P
C3	997813671	125	0.45	0.05	0.25	0.2	3.62	1.74	1.55	0.25	0.08	NM
A3	205325263	125	0.5	0.15	0.25	0.2	27.65	3.88	23.26	0.36	0.15	NM
E4	37010091	50	0.4	0.15	0.25	0.2	44.69	14.87	29.31	0.31	0.21	P
C4	1029598188	125	0.4	0.1	0.25	0.2	40.07	15.65	23.85	0.35	0.21	P
A4	234254185	75	0.6	0.1	0.35	0.2	13.56	3.11	9.93	0.3	0.22	P
C5	1055100420	125	0.6	0.1	0.35	0.2	18.43	5.07	12.84	0.33	0.2	P
E7	78082670	125	0.5	0.05	0.35	0.2	1.95	0.45	1.22	0.23	0.03	NM
A6	277013841	50	0.4	0.05	0.35	0.2	3.54	1.59	1.6	0.22	0.13	P

Tabla A6: Valores de la función objetivo y cada término de ésta para todas las calibraciones lanzadas, siendo P=Progreso y NME=Número máximo de evaluaciones.