

Proyecto Fin de Carrera
Ingeniería de Civil

Aplicación de las técnicas de Machine Learning a la
predicción del fallo de tuberías en las redes de
abastecimiento de EMASESA

Autor: Francisco Jesús González Lozano

Tutor: Antonio Manuel Blázquez Gámez

José Cañas Delgado

Dpto. Elasticidad y Resistencia de Materiales
Escuela Técnica Superior de Ingeniería
Universidad de Sevilla

Sevilla, 2020



Proyecto Fin de Carrera
Ingeniería Civil

Aplicación de las técnicas de Machine Learning a la predicción del fallo de tuberías en las redes de abastecimiento de EMASESA

Autor:

Francisco Jesús González Lozano

Tutor:

Antonio Manuel Blázquez Gámez

Catedrático de Universidad

José Cañas Delgado

Catedrático de Universidad

Dpto. de Elasticidad y Resistencia de Materiales

Escuela Técnica Superior de Ingeniería

Universidad de Sevilla

Sevilla, 2020

Proyecto Fin de Carrera: Aplicación de las técnicas de Machine Learning a la predicción del fallo de tuberías
en las redes de abastecimiento de EMASESA

Autor: Francisco Jesús González Lozano

Tutor: Antonio Manuel Blázquez Gámez
José Cañas Delgado

El tribunal nombrado para juzgar el Proyecto arriba indicado, compuesto por los siguientes miembros:

Presidente:

Vocales:

Secretario:

Acuerdan otorgarle la calificación de:

Sevilla, 2020

El Secretario del Tribunal

Agradecimientos

A mis padres, José y María Pilar, y a mi hermano, José Domingo, por todo el apoyo y ánimo que me han transmitido en mi etapa académica. A mi padre, por haber sido mi ejemplo para comprender el significado de sacrificio y perseverancia, valores sin los cuales, nunca podría haber alcanzado el objetivo que me propuse. A mi madre, por darme en todo momento ese cariño tan fundamental y haberme ayudado a levantar la cabeza, en esos duros momentos cuando sentía que todo se volvía oscuro. A mi hermano, por haberse preocupado de mi desarrollo académico y por haberme hecho sonreír en aquellos momentos donde el estrés estaba más presente.

A María, que se ha vuelto un pilar fundamental en mi vida, por estar siempre a mi lado cuando más necesitaba un apoyo, por hacerme ver el mundo desde otros puntos de vista.

A mis amigos, especialmente a mi gran amigo Carlos, por poder contar con ellos, y por brindarme una mano siempre que la he necesitado.

A mis profesores, que me han enseñado todo, por haber despertado en mí el interés por aprender. Especialmente a mis tutores en este trabajo, Antonio, por haber trabajado a mi lado, por haberme transmitido sus conocimientos y consejos, y Pepe, por haberme permitido utilizar los Softwares en los que ha invertido tanto tiempo y esfuerzo.

Sin ellos, no hubiera sido posible alcanzar mi objetivo.

Llega a su fin mi primera etapa universitaria, de la que no podría estar más orgulloso.

Gracias a todos por haberme ayudado a ser quien soy y a cumplir mis objetivos.

Fran

Sevilla, septiembre de 2020

En la actualidad, existen gran variedad de campos de la Ingeniería que cada vez más están implementando la metodología *Machine Learning*, o Aprendizaje Automático.

Principalmente esta metodología se emplea en modelos que tienen como finalidad predicciones, aprendizaje supervisado o aprendizaje no supervisado.

Tras una investigación realizada por FIUS (Fundación de Investigación de la Universidad de Sevilla) se elaboró un procedimiento basado en Aprendizaje Automático con la finalidad de predecir el fallo de las tuberías pertenecientes a la red de abastecimiento de EMASESA (Empresa Metropolitana de Abastecimiento y Saneamiento de Aguas de Sevilla), mediante un modelo de regresión lineal multiparámetro, que tiene como finalidad la asignación de una puntuación a cada tubería mediante el parámetro PF, el cual asigna un cierto riesgo de incidencia.

En este Trabajo Fin de Grado se emplea esta metodología con el fin de presentar resultados que garanticen la relación entre el factor PF con la Probabilidad de Fallo real de las tuberías, así como un estudio exhaustivo de los datos usados en el modelo de regresión, para comprobar su calidad.

Se aplican, además, una serie de procedimientos estadísticos basados en normalización y en tablas de contingencia, con el fin de optimizar la calidad de los datos para la obtención de unos mejores resultados, y comprobar las posibles correlaciones entre los factores que definen los datos de entrada o *inputs*.

Abstract

At present, there are a great variety of fields of Engineering that are increasingly implementing the Machine Learning methodology.

Mainly this methodology is used in models whose purpose is Predictions, Supervised Learning or Unsupervised Learning.

After an investigation carried out by FIUS (Research Foundation of the University of Seville), a procedure based on Machine Learning was developed in order to predict the failure of the pipes belonging to the supply network of EMASESA (Metropolitan Company of Supply and Sanitation of Seville's Waters), using a multiparameter linear regression model, whose purpose is to assign a score to each pipeline using the PF parameter, which assigns a certain risk of incidence.

In this Final Degree Project, this methodology is used in order to present results that guarantee the relationship between the PF factor and the Probability of real failure of the pipes, as well as an exhaustive study of the data used in the regression model, to check its quality.

In addition, a series of statistical procedures based on normalization and contingency tables will be carried out in order to optimize the quality of the data to obtain better results, and verify the possible existence between the factors that define the input data. or inputs.

Agradecimientos	vii
Resumen	ix
Abstract	xi
Índice	xiii
Índice de Tablas	xv
Índice de Figuras	xix
1 Introducción	1
1.1 <i>Introducción y Antecedentes</i>	1
1.2 <i>Descripción general de las redes de abastecimiento de EMASESA</i>	4
1.2.1 Descripción general de la red	5
1.2.2 Fuentes de Datos	5
1.3 <i>Modelo desarrollado. Método de regresión lineal multiPAámetro</i>	7
1.3.1 Definición de la solución Implementada.	7
1.3.2 Factores considerados.....	9
1.3.3 Conclusiones sobre el Uso de Datos desarrolladas por el GERM.	14
1.4 <i>Objetivo</i>	15
1.5 <i>Metodología</i>	16
1.6 <i>Contenido del trabajo</i>	16
1.7 <i>Descripción de software utilizado</i>	17
1.7.1 Software para Generación de Datos	17
1.7.2 Software para Análisis de Factores	18
1.7.3 Software de Ajuste y Predicción	20
2 Estudio de Factores.....	23
2.1 <i>Análisis de Factores Cualitativos</i>	23
2.1.1 Tablas de Contingencia.	24
2.1.2 Desarrollo del Proceso y Resultados.	26
2.2 <i>Efecto de los factores sobre la aparición de incidencias</i>	30
2.2.1 Análisis de distribución de los factores.....	30
2.2.2 Análisis del efecto de los factores sobre la aparición de incidencias.....	41
2.2.3 Conclusiones sobre el Análisis de Factores. Filtrado del número de registros.	52
2.3 <i>Análisis del efecto de filtrado de Outliers. Comparación de Resultados</i>	54
3 Tratamiento de Datos y Contraste de Resultados.....	60
3.1 <i>Análisis de Factores Cuantitativos</i>	60
3.1.1 Justificación del proceso de Normalización.....	60
3.1.2 Proceso de Normalización.	61
3.1.3 Estudio de Factores Susceptibles de Normalización.....	62
3.1.4 Resultados por Quinquenios. Contraste de Resultados.	64
3.1.5 Análisis de Coeficientes de Ajuste. Estudio de Pesos.....	67
3.2 <i>Constatación de relación entre PF y Probabilidad de Fallo</i>	78
3.2.1 Descripción del Desarrollo	78
3.2.2 Resultados.....	79

4	Conclusiones y Desarrollos Futuros.....	84
4.1	<i>Conclusiones del Capítulo 2. Análisis de Factores</i>	<i>84</i>
4.1.1	Conclusiones sobre la Correlación de Factores Cualitativos.....	84
4.1.2	Conclusiones sobre filtrado de <i>Outliers</i>	84
4.2	<i>Conclusiones del Capítulo 3. Tratamiento de Datos y Contraste de Resultados.....</i>	<i>85</i>
4.2.1	Conclusiones Acerca del Tratamiento de Datos.	85
4.2.2	Conclusiones sobre la Comparación Factor PF-Probabilidad de Fallo.	86
4.3	<i>Desarrollos Futuros</i>	<i>86</i>
4.3.1	Correlación de Factores Cualitativos	86
4.3.2	Filtrado de <i>Outliers</i>	86
4.3.3	Procedimiento de normalización y los coeficientes de ajuste.....	86
4.3.4	Relación del índice PF con la Probabilidad de Fallo	86
5	Bibliografía.....	87
6	Anexo. Tablas de Resultados	89
6.1	<i>Tablas de Resultados Estudio de Factores.....</i>	<i>89</i>
6.2	<i>Resultados Análisis efecto de filtrado de Outliers. Matrices de Confusión.</i>	<i>104</i>
6.2.1	Análisis de Bondad por quinquenios con Datos Filtrados. Resultados.	104
6.2.2	Análisis de Bondad por quinquenios con Datos Originales. Resultados.	107
6.3	<i>Resultados Análisis de Factores Cuantitativos. Matrices de Confusión.....</i>	<i>111</i>
6.3.1	Análisis de bondad por quinquenios. Datos normalizados.....	111
6.3.2	Resultados de análisis de bondad para datos normalizados.	114

ÍNDICE DE TABLAS

Tabla 1. Estructura de datos obtenida a través de GIS y AQUAWS-RED. Ejemplo.	6
Tabla 2. Estructura de datos histórico de incidencias. Ejemplo.	6
Tabla 3. Matriz de Confusión. Estructura.	8
Tabla 4. Frecuencias Observadas. Factores Material y Ubicación.	26
Tabla 5. Frecuencias Esperadas. Factores Material y Ubicación.	26
Tabla 6. Términos Sumatorio Coeficiente de Contingencia Chi cuadrado. Factores Material y Ubicación.	27
Tabla 7. Frecuencias Observadas. Factores Material y Distrito.	27
Tabla 8. Frecuencias Esperadas. Factores Material y Distrito.	28
Tabla 9. Términos Sumatorio Coeficiente de Contingencia Chi cuadrado. Factores Material y Distrito.	28
Tabla 10. Frecuencias Observadas. Factores Ubicación y Distrito.	29
Tabla 11. Frecuencias Esperadas. Factores Ubicación y Distrito.	29
Tabla 12. Términos Sumatorio Coeficiente de Contingencia Chi cuadrado. Factores Ubicación y Distrito.	29
Tabla 13. Longitud frente a Tipo de Red y Material.	30
Tabla 14. Longitud frente a Material y Municipio.	31
Tabla 15. Longitud por Tipo de Red y Diámetro.	33
Tabla 16. Longitud frente a Tipo de Red y Ubicación.	34
Tabla 17. Longitud frente a Ubicación y Municipio.	35
Tabla 18. Longitud frente a Tipo de Red y Distrito.	36
Tabla 19. Longitud frente a Tipo de Red y Año de Instalación.	37
Tabla 20. Longitud frente a Tipo de Red y Municipio.	41
Tabla 21. Incidencias absolutas e incidencias por unidad de longitud en función del factor Material.	42
Tabla 22. Incidencias absolutas e incidencias por unidad de longitud en función del factor Diámetro.	43
Tabla 23. Incidencias acumuladas y por unidad de longitud frente al factor Ubicación.	45
Tabla 24. Incidencias acumuladas y por unidad de longitud frente a Distrito.	46
Tabla 25. Incidencias acumuladas y incidencias por unidad de longitud frente a año de Instalación.	47
Tabla 26. Procesos de filtrado realizados.	54
Tabla 27. Resumen resultados Datos Simplificados vs Datos Generales.	55
Tabla 28. Promedios Resultados Simplificados vs Globales.	55
Tabla 29. Valores De promedio y desviación estándar para Normalización.	63
Tabla 30. Datos Originales.	63
Tabla 31. Datos Normalizados.	64
Tabla 32. Resumen Resultados por quinquenios Datos Normalizados vs Sin Normalizar.	65
Tabla 33. Promedio Resultados por quinquenios Datos Normalizados vs Sin Normalizar.	65
Tabla 34. Coeficientes de Ajuste para Factores Cuantitativos Normalizados.	68
Tabla 35. Coeficientes de Ajuste para valores del factor cualitativo Material.	68
Tabla 36. Coeficientes de Ajuste para valores del factor cualitativo Ubicación.	69
Tabla 37. Coeficientes de Ajuste para valores del factor cualitativo Distrito.	70
Tabla 38. Coeficientes para comparación de Factores Cuantitativos y Cualitativos normalizados.	76
Tabla 39. Estructura de Resultados para el factor PF.	79
Tabla 40. Probabilidad de Fallo frente a Previsión de Fallo. Período de Ajuste 2003-2010. Predicción 2011.	79
Tabla 41. Probabilidad de Fallo frente a Previsión de Fallo. Período de Ajuste 2008-2015. Predicción 2017.	80

Tabla 42. Probabilidad de Fallo frente a Previsión de Fallo. Período de Ajuste 2010-2017. Predicción 2018.	82
Tabla 43. Probabilidad de Fallo frente a Previsión de Fallo. Período de Ajuste 2003-2017. Predicción 2018.	83
Tabla 44. Longitud de Red frente a Distritos.	90
Tabla 45. Longitud en función de Número de Acometidas. Familias	91
Tabla 46. Incidencias acumuladas e Incidencias por unidad de longitud frente a Número de acometidas. Familias.	93
Tabla 47. Incidencias acumuladas e incidencias por unidad de longitud frente a Número de Acometidas. Tramos.	94
Tabla 48. Longitud de RSA frente a Presión Media.	96
Tabla 49. Longitud de RAA frente a Presión Media.	99
Tabla 50. Incidencias e Incidencias por Ud de longitud frente a Presión Media.	101
Tabla 51. Longitud RSA frente a Fluctuación de Presión.	102
Tabla 52. Longitud de RAA frente a Fluctuación de Presión.	102
Tabla 53. Incidencias Acumuladas e Incidencias por Ud de Longitud frente a Fluctuación de Presión.	103
Tabla 54. Resultados por familias para Datos Simplificados. Quinquenio 2003-2008.	104
Tabla 55. Resultados por familias para Datos Simplificados. Quinquenio 2004-2009.	104
Tabla 56. Resultados por familias para Datos Simplificados. Quinquenio 2005-2010.	104
Tabla 57. Resultados por familias para Datos Simplificados. Quinquenio 2006-2011.	105
Tabla 58. Resultados por familias para Datos Simplificados. Quinquenio 2007-2012.	105
Tabla 59. Resultados por familias para Datos Simplificados. Quinquenio 2008-2013.	105
Tabla 60. Resultados por familias para Datos Simplificados. Quinquenio 2009-2014.	106
Tabla 61. Resultados por familias para Datos Simplificados. Quinquenio 2010-2015.	106
Tabla 62. Resultados por familias para Datos Simplificados. Quinquenio 2011-2016.	107
Tabla 63. Resultados por familias para Datos Simplificados. Quinquenio 2012-2017.	107
Tabla 64. Resultados por familias para Datos Originales. Quinquenio 2003-2008.	107
Tabla 65. Resultados por familias para Datos Originales. Quinquenio 2004-2009.	107
Tabla 66. Resultados por familias para Datos Originales. Quinquenio 2005-2010.	108
Tabla 67. Resultados por familias para Datos Originales. Quinquenio 2006-2011.	108
Tabla 68. Resultados por familias para Datos Originales. Quinquenio 2007-2012.	109
Tabla 69. Resultados por familias para Datos Originales. Quinquenio 2008-2013.	109
Tabla 70. Resultados por familias para Datos Originales. Quinquenio 2009-2014.	109
Tabla 71. Resultados por familias para Datos Originales. Quinquenio 2010-2015.	109
Tabla 72. Resultados por familias para Datos Originales. Quinquenio 2011-2016.	110
Tabla 73. Resultados por familias para Datos Originales. Quinquenio 2012-2017.	110
Tabla 74. Resultados análisis de bondad datos originales. Quinquenio 2003-2008.	111
Tabla 75. Resultados análisis de bondad datos originales. Quinquenio 2004-2009.	111
Tabla 76. Resultados análisis de bondad datos originales. Quinquenio 2005-2010.	111
Tabla 77. Resultados análisis de bondad datos originales. Quinquenio 2006-2011.	112
Tabla 78. Resultados análisis de bondad datos originales. Quinquenio 2007-2012.	112
Tabla 79. Resultados análisis de bondad datos originales. Quinquenio 2008-2013.	112
Tabla 80. Resultados análisis de bondad datos originales. Quinquenio 2009-2014.	113
Tabla 81. Resultados análisis de bondad datos originales. Quinquenio 2010-2015.	113
Tabla 82. Resultados análisis de bondad datos originales. Quinquenio 2011-2016.	113
Tabla 83. Resultados análisis de bondad datos originales. Quinquenio 2012-2017.	114
Tabla 84. Resultados análisis de bondad datos normalizados. Quinquenio 2003-2008.	114
Tabla 85. Resultados análisis de bondad datos normalizados. Quinquenio 2004-2009.	114
Tabla 86. Resultados análisis de bondad datos normalizados. Quinquenio 2005-2010.	115

Tabla 87. Resultados análisis de bondad datos normalizados. Quinquenio 2006-2011.	115
Tabla 88. Resultados análisis de bondad datos normalizados. Quinquenio 2007-2012.	115
Tabla 89. Resultados análisis de bondad datos normalizados. Quinquenio 2008-2013.	116
Tabla 90. Resultados análisis de bondad datos normalizados. Quinquenio 2009-2014.	116
Tabla 91. Resultados análisis de bondad datos normalizados. Quinquenio 2010-2015.	116
Tabla 92. Resultados análisis de bondad datos normalizados. Quinquenio 2011-2016.	117
Tabla 93. Resultados análisis de bondad datos normalizados. Quinquenio 2012-2017.	117

ÍNDICE DE FIGURAS

Figura 1 Acueducto de Les Ferreres (Tarragona).	1
Figura 2. Acequia del Turia.	1
Figura 3. Obras de Reparación de la Red de Abastecimiento de Eivissa.	2
Figura 4. Reparación de tubería de Abastecimiento (Montilla).	2
Figura 5. Evolución de pérdidas reales de agua en redes de distribución por fugas y averías (Fuente: INE).	3
Figura 6. Interfaz software Generación de Datos.	17
Figura 7. Interfaz software Análisis de factores. Análisis de la Red.	18
Figura 8. Interfaz software Análisis de Factores. Análisis de Incidencias.	19
Figura 9. Interfaz software Análisis de Factores. Efecto de las Incidencias Previas.	19
Figura 10. Interfaz. software de Análisis de Factores. 1Factor.	19
Figura 11. Software de Análisis de Factores. 2Factores.	20
Figura 12. Interfaz del Software de Ajuste y Predicción.	20
Figura 13. Software de Ajuste y Predicción. Definición de Factores para Análisis.	21
Figura 14. Software de Ajuste y Predicción. Opciones de Escritura.	21
Figura 15. Software de Ajuste y Predicción. Valores para PF.	21
Figura 16. Estructura Tabla de contingencia	24
Figura 17. Longitud frente a Tipo de Red y Material.	31
Figura 18. Longitud frente a Material y Municipio.	31
Figura 19. Longitud por Tipo de Red y Diámetro.	33
Figura 20. Longitud frente a Tipo de Red y Ubicación.	34
Figura 21. Longitud frente a Ubicación y Municipio.	35
Figura 22. Longitud frente a Distrito.	36
Figura 23. Longitud frente a Distrito y Tipo de Red.	37
Figura 24. Longitud frente a Tipo de Red y Año de Instalación.	38
Figura 25. Longitud frente a Año de Instalación. Años de Instalación más recientes.	38
Figura 26. Longitud frente a factor Número de Acometidas y Tipo de Red. Familias.	39
Figura 27. Número de Acometidas por unidad de longitud correspondiente a familias.	39
Figura 28. Longitud frente a factores Presión media y Tipo de Red.	40
Figura 29. Longitud frente a Fluctuacion de Presion y Tipo de Red.	41
Figura 30. Longitud red frente a Tipo de Red y Municipio.	42
Figura 31. Incidencias acumuladas por Material.	42
Figura 32. Incidencias por km para factor Material.	43
Figura 33. Incidencias acumuladas por Diámetro.	44
Figura 34. Incidencias por km para factor Diámetro.	44
Figura 35. Incidencias acumuladas por Ubicación.	45
Figura 36. Incidencias por km para factor Ubicación.	45
Figura 37. Incidencias por Distrito.	46
Figura 38. Incidencias por km para factor Distrito.	46
Figura 39. Incidencias acumuladas por Año de Instalación	48
Figura 40. Incidencias por km para factor Año de Instalación.	48
Figura 41. Incidencias acumuladas por Número de Acometidas. Familias.	48

Figura 42. Incidencias por km para factor Número de Acometidas. Familias.	49
Figura 43. Incidencias acumuladas por Número de acometidas. Tramos.	49
Figura 44. Incidencias por km para factor Número de Acometidas. Tramos.	50
Figura 45. Incidencias acumuladas totales por Presión Media.	50
Figura 46. Incidencias por km para factor Presión Media.	51
Figura 47. Incidencias acumuladas frente a Fluctuación de Presión.	51
Figura 48. Incidencias por unidad de longitud frente a Fluctuación de Presión.	52
Figura 49. Sensibilidad Datos Simplificados vs Datos Originales.	56
Figura 50. Valor Predictivo Positivo Datos Simplificados vs Datos Globales.	57
Figura 51. Índice de calidad para datos Simplificados vs Datos Globales.	57
Figura 52. Comparación Valores Medios Datos Simplificados vs Datos Globales.	58
Figura 53. Estructura de datos normalizados en Modelo de Ajuste y Predicción.	64
Figura 54. Sensibilidad para Datos normalizados vs Datos sin normalizar.	66
Figura 55. Valor Predictivo Positivo para Datos normalizados vs Datos sin normalizar.	66
Figura 56. Índice de Calidad para Datos normalizados vs Datos sin normalizar.	67
Figura 57. Evolución de coeficientes de Ajuste Normalizados para Factores Cuantitativos Normalizados. Con IPA.	71
Figura 58. Evolución de coeficientes de Ajuste Normalizados para Factores Cuantitativos Normalizados. Sin IPA.	71
Figura 59. Evolución de Coeficientes de Ajuste No Normalizados para Factor Cuantitativo Número de Incidencias Previas.	72
Figura 60. Evolución de Coeficientes de Ajuste No Normalizados para Factor Cualitativo Material. Con IPA.	73
Figura 61. Evolución de Coeficientes de Ajuste No Normalizados para Factor Cualitativo Material. Sin IPA.	74
Figura 62. Evolución de coeficientes de Ajuste Normalizados para Factor Cualitativo Ubicación. Con IPA.	74
Figura 63 . Evolución de coeficientes de Ajuste Normalizados para Factor Cualitativo Ubicación. Sin IPA.	75
Figura 64. Evolución de coeficientes de Ajuste Normalizados para Factor Cualitativo Distrito. Con IPA.	75
Figura 65. Evolución de coeficientes de Ajuste Normalizados para Factor Cualitativo Distrito. Sin IPA.	76
Figura 66. Relación entre coeficientes de ajuste a nivel General. Con IPA.	77
Figura 67. Relación entre coeficientes de ajuste a nivel General. Con IPA.	77
Figura 68. Relación factor PF frente a Probabilidad de Fallo. Período de Ajuste 2003-2010. Predicción 2011.	80
Figura 69. Relación factor PF frente a Probabilidad de Fallo. Período de Ajuste 2008-2015. Predicción 2017.	81
Figura 70. Relación factor PF frente a Probabilidad de Fallo. Período de Ajuste 2010-2017. Predicción 2018.	82
Figura 71. Relación factor PF frente a Probabilidad de Fallo. Período de Ajuste 2003-2017. Predicción 2018.	83

LISTADO DE SIGLAS

- RAA: Red Arterial y de Transporte
- RSA: Red Secundaria de Abastecimiento
- GIS: Geographic Information System
- BI: Business Intelligence
- ID: Identificador tramo GIS
- PF: Índice de Previsión de Fallo
- NIT: Número de Incidencias Totales
- LI: Límite Inferior
- LS: Límite Superior
- VP: Verdaderos Positivos
- VN: Verdaderos Negativos
- FP: Falsos Positivos
- FN: Falsos Negativos
- IC: Índice de Calidad
- FC: Fibrocemento
- HF: Hierro Fundido
- FD: Fundición dúctil
- HACH: Hormigón Armado con Chapa
- PE: Polietileno
- PE BD/MD: Polietileno Baja Densidad/Media Densidad
- PVC: Policloruro de Vinilo
- AI: Acero Inoxidable
- BBDD: Bases de Datos
- IPA: Incidencias Previas Acumuladas
- ML: Machine Learning

1 INTRODUCCIÓN

1.1 Introducción y Antecedentes.

La necesidad de generar y sostener infraestructuras para el tratamiento y distribución de aguas en los núcleos poblacionales ha sido a lo largo de la historia, uno de los problemas más importantes de la Ingeniería Civil, dada la importancia de disponer de agua potable de forma continua y cómoda, y evacuar las aguas residuales y pluviales.

Los primeros sistemas de abastecimiento y saneamiento de agua datan de la época romana, donde se diseñaron las primeras cloacas para el transporte de aguas negras, o acueductos (Figura 1, descargada de: <https://viajandoporviajar.blogspot.com/2018/06/acueducto-romano-de-tarragona-el-puente.html>, el 03/08/2020), para la distribución de agua potable.

También, otras civilizaciones como la musulmana, mediante las acequias (Figura 2, descargada de: <https://www.levante-emv.com/horta/2016/07/04/acequias-corazon-arterias-lhorta/1440247.html>, el 03/08/2020), o la China, mediante estructuras de bambú, satisfacían estas necesidades.



Figura 1. Acueducto de Les Ferreres (Tarragona).



Figura 2. Acequia del Turia.

Los sistemas de abastecimiento y saneamiento modernos se comienzan a desarrollar a partir del siglo XVIII. En ese momento comienzan a emplearse materiales metálicos como hierro o cobre, quedando obsoletos materiales como la piedra o arcilla, que se usaban anteriormente. Esta evolución continúa hasta el siglo XX donde aparecen materiales como el Fibrocemento, el hormigón, los materiales plásticos, etc.

En el contexto de Sevilla, en 1968 nació el Servicio Municipal de Aguas y ya en 1974 la Empresa Municipal de Abastecimiento y Saneamiento de Aguas de Sevilla, hoy en día conocida como

EMASESA (fuente: <https://www.emasesa.com/emasesa-muestra-la-historia-del-abastecimiento-de-agua-de-sevilla-en-fotografias/>).

En general, los sistemas han ido evolucionando, adaptándose a las necesidades que iban surgiendo, como consecuencia de la expansión de las ciudades y de la creciente demanda debido al gran aumento de la densidad de población.

El avance en el campo de la Ingeniería Hidráulica y la Mecánica de Fluidos ha permitido dotar a las ciudades de grandes redes que abarcan kilómetros de longitud, y que permiten abastecer de agua potable directamente en el núcleo residencial, diseñando sistemas subterráneos de conducciones que evitan posibles interferencias con el movimiento cotidiano en la superficie.

No obstante, no es solo el diseño de la infraestructura el foco de estudio, sino que un pilar fundamental con la misma o incluso más relevancia que éste, es la conservación y el mantenimiento que requieren estos sistemas, dada la gran importancia de dotar tanto de cantidad como de calidad de agua suficiente a la población.

Con el paso del tiempo, estos sistemas están sometidos a las inclemencias externas: corrosión, sobrecargas debido a actuaciones cercanas, movimientos del terreno, tráfico en superficie, etc. que hacen que vayan envejeciendo y deteriorándose, generando así fugas que pueden dar lugar a la pérdida de una parte importante del agua que se distribuye.

Por otra parte, la necesidad de realizar una reparación o una sustitución en una red de abastecimiento conlleva un coste bastante elevado tanto en términos económicos como en términos logísticos (Figura 3, descargada de: <https://www.periodicodeibiza.es/pitiusas/ibiza/2014/08/16/131024/eivissa-pierde-000-millones-litros-agua-potable-ano-por-red-suministro-urbano.html>, el 01/08/2020; Figura 4, descargada de: <https://www.aguasdemontilla.es/-/estado-de-las-obras-de-reparacion-de-la-tuberia-principal-de-abastecimiento-a-montil-1>, el 01/08/2020).



Figura 3. Obras de Reparación de la Red de Abastecimiento de Eivissa.



Figura 4. Reparación de tubería de Abastecimiento (Montilla).

Se muestra en la figura 5, una representación gráfica de datos oficiales extraídos del INE (Instituto Nacional de Estadística), con la evolución a nivel nacional, y para la comunidad autónoma de Andalucía, del volumen de agua perdida en la red de distribución como consecuencia de fugas y averías (datos hasta 2016), en litros por habitante.

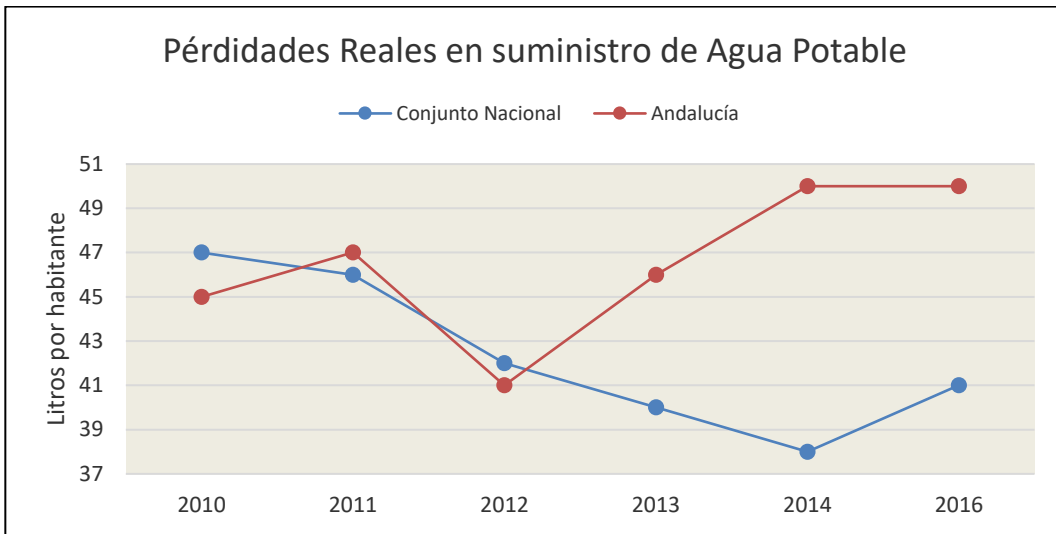


Figura 5. Evolución de pérdidas reales de agua en redes de distribución por fugas y averías (Fuente: INE).

Las pérdidas en Andalucía han sido de 50 litros por habitante al día en los últimos años consultados, aproximadamente un 20% superiores a la media española, es decir, en Andalucía, entre 2014 y 2016, las fugas y las roturas, o en general “incidencias” en las redes de abastecimiento supusieron una pérdida de casi la mitad del consumo de la población.

Según el INE, el consumo medio por habitante en España es de 132 litros diarios, por lo que las pérdidas (más del 40% del consumo medio) suponen un problema muy importante dentro del campo de la gestión de los sistemas de abastecimiento.

Si se calcula el producto de este consumo medio por el 0.41 litros que se pierden por cada litro de consumo, se obtiene que el volumen de agua perdido al día por habitante alcanza un valor de 54.12 litros. Extrapolando este resultado al período anual se obtiene que al año se producen unas pérdidas de hasta 19753.8 litros por habitante.

En España, existen en torno a 40 millones de habitantes, por lo que las pérdidas anuales en el territorio nacional equivaldrían a 790e9 litros perdidos al año. Esta cantidad equivale a 790 Hm³ de pérdidas por año.

Los pantanos que abastecen el área metropolitana de Sevilla son: Aracena (129 Hm³) y Zufre (175 Hm³), situados en la provincia de Huelva, y La Minilla (58 Hm³), El Gergal (35 Hm³), Los Melonares (186 Hm³) y Cala (59 Hm³), en la provincia de Sevilla. La suma del volumen de todos los pantanos supone más de 600 Hm³ (<https://www.embalses.net/provincia-6-sevilla.html>), es decir, no sería suficiente para compensar las pérdidas de todo el territorio nacional.

Otro ejemplo de la magnitud de estas pérdidas fue que tras un estudio sobre la red de abastecimiento en la comunidad gallega realizado por AEAS (Asociación Española de Abastecimiento y Saneamiento de Agua), se determinó que una cuarta parte del agua potable (equivalente a 40 hectómetros cúbicos anuales) se pierde en las redes de abastecimiento, principalmente por averías y fugas en determinados puntos de la red. En ese mismo estudio se concluye que en un funcionamiento óptimo de una red de abastecimiento, las pérdidas no deberían superar el 10%. (fuente: https://www.abc.es/espana/galicia/abci-cuarta-parte-agua-potable-pierde-redes-abastecimiento-201708270857_noticia.html).

Esta conclusión se da en la mayoría de las redes de abastecimiento en grandes núcleos poblacionales, donde la cantidad de agua demandada alcanza niveles muy elevados.

Tras observar esta problemática, aparece la necesidad de analizar el estado de las redes de abastecimiento de forma eficiente. Por ello surge el interés de las empresas suministradoras en disponer de herramientas que les permitan controlar la calidad de la red que gestionan y anticIPAs a las posibles situaciones conflictivas con medidas proactivas.

Aun así, no existen en la actualidad herramientas para tales fines a nivel universal, sino que se desarrollan de forma personalizada para cada red particular. Aunque las técnicas sean las mismas (en general), las conclusiones no son directamente extrapolables entre la situación de la red de una ciudad y la de otra.

Una de las técnicas más populares en la actualidad fundamenta la herramienta en el análisis estadístico de los datos históricos, empleando para ello las metodologías de Aprendizaje Automático o *Machine Learning*.

En línea con estas inquietudes a nivel general en los países desarrollados, EMASESA ha liderado un proyecto de investigación que plantea la necesidad de desarrollar un modelo predictivo con la finalidad de obtener una estimación del número de incidencias que habría en años posteriores, a corto y medio plazo en las redes de abastecimiento y saneamiento, en el entorno del área metropolitana de Sevilla.

En el documento [1] se realizó un estudio del estado de arte, donde se analizaron distintos modelos propuestos por investigadores, científicos e ingenieros en situaciones similares. Este estudio abarca desde modelos deterministas, estadístico-deterministas, estadístico-probabilistas, físico-probabilistas, modelos de redes neuronales, modelos basados en lógica Fuzzy, y modelos heurísticos.

Cada uno de estos presentaba particularidades propias. Se va a comentar de forma breve una clasificación de los modelos que se estudiaron, comentando cada clase o grupo de ellos.

- **Con base física.** Intentan predecir el comportamiento de la red resolviendo el problema del fenómeno físico asociado a la incidencia (corrosión, rotura, etc). Tienen el inconveniente de que necesitan conocer con suficiente precisión todos los datos (propiedades y cargas) que condicionan el fenómeno físico concreto de que se trate, conocer todas las situaciones que solicitarán al sistema a lo largo de su vida útil y disponer de un modelo matemático suficientemente fiable. Tienen la ventaja de que suelen ser procedimientos muy precisos para determinar la respuesta del sistema ante unas situaciones determinadas.
- **Con base estadística.** Se trata de analizar los históricos de incidencias de la red con las herramientas del análisis estadístico, incluyendo las técnicas de Machine Learning. Tienen la ventaja de que no necesitan tanta información como los modelos de base física, y que pueden funcionar con los datos de que se disponga. Tienen el inconveniente de ser muy sensibles a la calidad y cantidad de los datos.
- **Heurísticos.** Son procedimientos que aprovechan la experiencia de los técnicos involucrados en el desarrollo del procedimiento. Son fácilmente adaptables para nuevas situaciones. Tienen el inconveniente de ser muy subjetivos y estar continuamente sometidos a críticas que cuestionan su validez.

1.2 Descripción general de las redes de abastecimiento de EMASESA.

En este apartado se realizará una breve introducción a la estructura de la red objeto de estudio, y se comentarán las distintas fuentes con las que se contó para el desarrollo de la investigación, con el objetivo de obtener los datos con los que posteriormente se trabajaría.

Para el desarrollo del TFG se ha tomado como referencia y base principal los documentos elaborados por GERM:

- [1] Análisis del estado del arte y asimilación por parte del adjudicatario de la información, conocimiento y herramientas desarrolladas por EMASESA.

- [2] Herramienta de priorización de actuaciones sobre una red de distribución de aguas basada en el índice de fallo (PARDIF). Aplicación a la red de abastecimiento.
- [3] Herramienta de priorización de actuaciones sobre una red de distribución de aguas basada en el índice de fallo (PARDIF). Manual de Usuario.

Los documentos mencionados pertenecen al Proyecto de investigación “**Investigación aplicada al análisis del comportamiento y durabilidad de los materiales de las redes de abastecimiento y saneamiento para la optimización de las herramientas de Gestión Patrimonial de Infraestructuras de EMASESA**”, en el cual pude participar mediante la obtención de una beca en competencia, de 9 meses de duración, para realizar algunas de las labores dentro del proyecto, tales como el aprendizaje sobre el uso de las bases de datos de EMASESA de las que posteriormente se hablará, aprendizaje sobre el manejo de los distintos Softwares desarrollados para conseguir el objetivo del proyecto, y estudio de varios documentos técnicos/tesis para el desarrollo de la investigación, entre otros.

1.2.1 Descripción general de la red

La Red de abastecimiento de EMASESA perteneciente al entorno del área metropolitana de Sevilla, se clasifica en dos subredes principales, la RAA (Red Arterial y de Transporte), y la RSA (Red Secundaria de Abastecimiento).

Según los datos que se obtuvieron, contabilizados hasta finales de 2018, se observó que la longitud total de la red de abastecimiento de EMASESA es de 3927.4 km, perteneciendo 487.9 km a la RAA y 3439.47 km a la RSA.

En cuanto al número de tramos, hay registrados un total de 112653 tramos hasta 2018, de los cuales 5224 pertenecen a la RAA y 107429 tramos pertenecen a la RSA.

La distribución de esta red se extiende a los municipios de Alcalá de Guadaíra, Alcalá del Río, Camas, Carmona, Coria del Río, Dos Hermanas, El Garrobo, El Ronquillo, Gelves, Guillena, La Puebla del Río, La Rinconada, Los Palacios y Villafranca, Mairena del Alcor, Palomares del Río, San Juan de Aznalfarache, Santiponce, Sevilla y Tomares.

1.2.2 Fuentes de Datos

EMASESA tiene organizada toda la información referente a la red de tuberías de abastecimiento y saneamiento mediante dos plataformas concretas: AQUAWS-RED y GIS, gestionadas por BI.

- GIS (Geographic Information System) es un marco para recopilar, ajustar y analizar datos, que agrupa información de todos los elementos que conforman las redes tanto de abastecimiento como de saneamiento, incluyendo información relativa a las características de cada tramo, tales como diámetro, longitud, material, etc. Al tratarse de un programa de Sistemas de Información Geográfica, la finalidad de este es representar en un entorno gráfico la distribución de los distintos componentes que conforman la red.
- La Plataforma AQUAWS-red, es un programa que gestiona la base de datos y que incluye información sobre el estado de la red en tiempo real. Incluye información relativa a los tramos, intervenciones que se han realizado sobre ellos, de los trabajadores que han realizado las intervenciones, y la relacionada con el mantenimiento de la red.
- Por otro lado, EMASESA implementa el procedimiento o metodología BI (Business Intelligence). Esta metodología se basa en el uso de datos en una empresa para facilitar la toma de decisiones. Gracias a esta metodología, se pueden realizar pedidos de grandes volúmenes de datos, provenientes de las plataformas anteriormente descritas para su posterior tratamiento.

El aprendizaje sobre el uso de las bases de datos mencionadas para obtener el volumen necesario de información fue uno de los papeles más relevantes que realicé en el desarrollo del Proyecto.

Los datos en estas plataformas están organizados en registros, correspondiendo a cada tramo un registro. La definición de un tramo es subjetiva, debido a que puede ser de gran longitud o de corta longitud,

aunque cada tramo está definido por un único material, diámetro, año de instalación, etc. A partir de ahora, el término tramo hará referencia a un tramo tal y como se entiende en las bases de datos de EMASESA.

Cada tramo tiene asignado un registro, que se compone por una serie de factores los cuales describen características intrínsecas al tramo, o de su entorno. Los factores tienen diferente naturaleza. Estos pueden ser:

- Factores Cuantitativos
- Factores Cualitativos
- Factores de Identificación

A partir de estas dos bases de datos (AQUAWS-red y GIS), un programa extrae todas las características de los tramos y los datos históricos sobre las incidencias ocurridas, y las almacena en dos hojas de EXCEL.

1. La primera hoja, que se puede observar en la tabla 1, con toda la información relativa a los tramos (registros).

Código GIS (Tramo Ab.)	Material	Diámetro nominal	Año Instalación	Longitud	Número de acometidas	Nivel de servicio	Puntuación Total	Tipo de red
10113185	HF	200	1950	2.085	0	Tramo corto, no aplica	0	Potable secundaria
10113254	HACH	800	1970	69.145	0	A SUSTITUIR - PRIORIDAD 1	-56.667	Potable transporte
10113325	FD	800	2006	12.501	0	BUENO	97.2	Potable transporte
10116199	FC	100	1980	6.884	0	MEDIO	67.778	Potable secundaria
10116663	FD	100	2001	24.017	1	BUENO	96.2	Potable secundaria
10346218	PE100	110	2006	2.195	0	Tramo corto, no aplica	0	Potable secundaria
10346256	PE100	110	2006	0.658	0	Tramo corto, no aplica	0	Potable secundaria
10346267	PE100	110	2006	2.449	0	Tramo corto, no aplica	0	Potable secundaria
10346431	PE100	110	2006	104.219	17	BUENO	93.4	Potable secundaria
10346442	PE100	110	2006	100.409	13	BUENO	93.4	Potable secundaria
10346554	PE100	160	2006	52.153	2	MEDIO	89.4	Potable secundaria
10346558	PE100	110	2006	0.35	0	Tramo corto, no aplica	0	Potable secundaria
10346567	PE100	110	2006	49.384	2	BUENO	94.4	Potable secundaria
10346576	PE100	110	2006	0.3	0	Tramo corto, no aplica	0	Potable secundaria
10346589	PE100	160	2006	4.167	0	BUENO	94.4	Potable secundaria
10346612	PE100	200	2006	0.275	0	Tramo corto, no aplica	0	Potable secundaria
10346664	PE100	160	2006	42.244	7	ÓPTIMO (100 > NS > 90)	96	Red secundaria
10346692	PE100	110	2006	0.25	0	Tramo corto, no aplica	0	Potable secundaria
10346706	PE100	110	2006	0.3	0	Tramo corto, no aplica	0	Potable secundaria
10346719	PE100	110	2006	0.817	0	Tramo corto, no aplica	0	Potable secundaria
...

Tabla 1. Estructura de datos obtenida a través de GIS y AQUAWS-RED. Ejemplo.

2. La segunda, cuya estructura aparece ilustrada en la tabla 2, contiene información acerca del número de incidencias que han aparecido en los tramos, a lo largo de su vida útil. La estructura de la hoja se dispone en dos columnas, una primera donde se define el ID (GIS), es decir, el código de identificación del tramo, y una segunda llamada “Año de actuación”, donde se define el año en el que ocurrió la incidencia.

Las dos hojas, a priori independientes, se relacionan entre ellas mediante el identificador “Código GIS”. Posteriormente, mediante un programa elaborado por el equipo técnico del GERM, implementado la Macro “Genera Datos” se obtendrá una hoja de Excel que relacionará las características de los tramos con las incidencias que hayan aparecido en estos, y con la fecha de aparición de cada incidencia.

ID (GIS)	Año Actuación
932856	2003
21285888	2003
14467833	2003
613147	2003
705653	2003
828601	2003
919321	2003
943168	2003
...	...

Tabla 2. Estructura de datos histórico de incidencias. Ejemplo.

1.3 Modelo desarrollado. Método de regresión lineal multiPAámetro.

1.3.1 Definición de la solución Implementada.

Finalmente se optó por el desarrollo de un modelo físico-probabilista con el objetivo de obtener una previsión de fallo en los distintos tramos que conforman la red, dada la imposibilidad de realizar un modelo físicamente basado debido a la complejidad del problema y a la cantidad de parámetros necesarios (muchos de ellos desconocidos).

Las características del modelo se describen en el documento [2]. El modelo se basa en plantear un problema de regresión multiPAámetro basado en mínimos cuadrados, con una variable de estimación, Y (número de incidencias esperable en el periodo de predicción), que se relaciona con una serie de factores, X_i (diámetro, material, ubicación, etc).

La idea principal de esta regresión es obtener unos coeficientes de ajuste a partir de un histórico de incidencias, que posteriormente serán aplicados al proceso de predicción. Una vez obtenidos los resultados del proceso de predicción en términos de número de incidencias, se implementa una operación basada en una función logística que permite trasladar el número de incidencias previsto a un parámetro denominado Previsión de Fallo, que cuantifica el riesgo entre 0 y 1. Obviamente, el interés de la herramienta está en la aplicación para estimar las incidencias en años futuros; pero en el caso de que la predicción se realice para un año pasado (para el que se dispone de datos históricos), la comparación entre las predicciones y los datos reales aportan evidencias sobre la fiabilidad de las predicciones para los años futuros.

La regresión que se dispone es de tipo lineal, que se representa mediante la expresión (1), donde N , es el número total de factores considerados, b_0 , es el término independiente, y b_i , los coeficientes de la correlación.

$$Y = b_0 + \sum_{i=1}^N b_i * X_i \quad (1)$$

El procedimiento general se estructura de la siguiente forma:

- 1a) De todos los datos se selecciona un periodo para entrenar el procedimiento, es decir se seleccionan los datos utilizados para ajustar los parámetros que definen el método. A este periodo se le llama periodo de ajuste o entrenamiento.
- 1b) Se selecciona un periodo para el que se realiza la estimación. Es el periodo de predicción o estimación. Si se dispone de datos reales, estos pueden utilizarse para comprobar la calidad de la predicción, y el periodo se llamaría también periodo de comprobación.
- 2) Después de esta selección, en el nivel global se predice (con el grado de confianza deseado) cuantas incidencias, en total, van a aparecer en la red en el periodo de estimación.
- 3) El ajuste se aplica a todos los componentes de la red (nivel local) obteniendo un nivel de riesgo de fallo (PF) para cada familia/tramo.
- 4) Finalmente, se ordenan por dicho índice y se clasifican según el número de incidencias totales predicho por el nivel global: bajo, medio, alto y muy alto. Además, se asigna el nivel de riesgo "Muy Alto" a aquellas NIT familias/tramos con mayor PF, siendo NIT el número de incidencias totales estimadas en el punto 2.

Así pues, existen dos niveles a la hora de ejecutar el modelo.

- 1) **Nivel global**, cuyo objetivo es determinar el número de incidencias totales esperadas en el periodo de predicción/comprobación con un cierto nivel de confianza.
- 2) **Nivel local**, cuyo objetivo es **predecir** el número de incidencias **previstas** para el periodo de predicción/comprobación.

A su vez, el procedimiento para el nivel local se divide en tres etapas, más 1 adicional de comprobación si se dispusieran de datos durante el periodo de estimación con los que comparar las predicciones:

- 1) **Etapa de Ajuste, entrenamiento o aprendizaje.** En esta etapa se parte del histórico de incidencias durante el período seleccionado, se procede a realizar el ajuste, siendo en este caso las incógnitas los coeficientes de correlación, y teniendo la variable Y , el valor de las incidencias acumuladas por cada uno de los tramos dentro del período de ajuste.
- 2) **Etapa de Predicción.** En esta etapa, una vez hallado el valor de los coeficientes de ajuste, se utilizan para obtener el número de incidencias que se estima pueden darse en el periodo seleccionado para hacer la predicción. El sistema de regresión lineal estaría formado por los coeficientes obtenidos anteriormente y los valores de los factores que definen los registros de los tramos de los que se quiere obtener el número de incidencias previstas.
- 3) **Etapa de Clasificación.** Tras haber calculado el número de incidencias previstas en el periodo de predicción considerado, se lleva a cabo el cálculo de la PF (Previsión de Fallo).

En general, el número de incidencias previsto con la regresión puede tomar cualquier valor entre $-\infty$ y $+\infty$, lo que lo hace difícil de manejar. Por ello se utiliza como predicción del fallo la función *logit* del número de incidencias previstas durante el periodo de estimación, que proporciona un resultado entre 0 y 1.

$$PF = \text{Logit}(\hat{y}) = \frac{1}{1 + \text{Exp}(-\hat{y})} \quad (2)$$

Dentro del nivel local, y una vez determinada esta previsión de fallo, se establecen unos límites para clasificar los tramos en tramos con previsión Muy Alta, Alta, Media o Baja. Estos límites son el Límite Superior (LS), y el Límite Inferior (LI).

Si el tramo posee una $PF \geq LS$ se le considera riesgo ALTO.

En el caso de que $PF \leq LI$, el riesgo asociado será BAJO.

Si $LI \leq PF \leq LS$, el riesgo asociado será medio.

El límite que separa los tramos clasificados como de riesgo MUY ALTO de los clasificados como riesgo ALTO, viene definido por el resultado del análisis global. [2]

Para realizar esta etapa se utiliza una herramienta fundamental, llamada “Matriz de Confusión”, [2].

- 4) **Etapa de Comprobación.** Esta etapa (adicional) consiste en contrastar los resultados obtenidos en la etapa de predicción (en el caso de que se dispongan de datos con los que comparar las predicciones) con el número de incidencias existentes realmente para el año de comprobación.

Esta matriz compara los valores de la estimación, y los valores reales que aparecen en el histórico.

La estructura de la matriz de confusión se muestra en la tabla 3.

Matriz de confusión		Estimación	
		Incidencias	No incidencias
Realidad	Incidencias	VP	FN
	No incidencias	FP	VN

Tabla 3. Matriz de Confusión. Estructura.

Los valores que aparecen representados en los cuadros de color azul claro y azul oscuro:

- VP (Verdaderos Positivos): este valor designa cuantos tramos para los que se han estimado incidencias (nivel de riesgo Muy Alto) realmente las han presentado.
- FN (Falsos Negativos): este valor designa cuantos tramos para los que no se han estimado incidencias (nivel de riesgo Muy Alto) realmente las han presentado.
- FP (Falsos Positivos): este valor designa cuantos tramos para los que se han estimado incidencias (nivel de riesgo Muy Alto) realmente no las han presentado.
- VN (Verdaderos Negativos): este valor designa cuantos tramos para los que no se han estimado incidencias (nivel de riesgo Muy Alto) realmente no las han presentado.

Una vez definidas las cuatro entradas que aparecen en la matriz de confusión, se establecen una serie de parámetros que se calculan a partir de estos valores, y que serán utilizados en la comparación entre las diferentes opciones contempladas en este TFG. Se definen a continuación.

- Sensibilidad: ratio entre los verdaderos positivos y el total de incidencias reales.

$$\text{Sensibilidad} = \frac{VP}{VP + FN} * 100$$

- Especificidad: ratio entre verdaderos negativos, y el total de no incidencias reales.

$$\text{Especificidad} = \frac{VN}{VN + FP} * 100$$

- Índice de validez: ratio entre total de verdaderos y total de tramos o familias.

$$\text{Índice de validez} = \frac{VP + VN}{VP + FP + FN + VN} * 100$$

- Valor predictivo (+): ratio entre verdaderos positivos y total de positivos:

$$\text{Valor predictivo (+)} = \frac{VP}{VP + FN} * 100$$

- Valor Predictivo (-): ratio entre verdaderos negativos y total de negativos.

$$\text{Valor predictivo (-)} = \frac{VN}{VN + FP} * 100$$

Estos valores permitirán comparar la calidad de las predicciones en los estudios realizados en este TFG.

Dado que en nuestro problema el número de tramos sin incidencias es mucho mayor que el número de tramos que sufren incidencias (es decir: $FN + VN \gg VP + FN$, y también $FN + VN \gg VP + FP$) y que el interés en nuestro problema es predecir con precisión qué tramos van a sufrir incidencias, los parámetros más interesantes son: la sensibilidad y el valor predictivo positivo. Por ello el parámetro que se va a utilizar principalmente es la semisuma de ambos, denominado “Índice de Calidad”

- Índice de Calidad:

Se define como la semisuma de la sensibilidad y el valor predictivo positivo.

$$IC = \frac{\text{Sensibilidad} + \text{Valor predictivo (+)}}{2}$$

1.3.2 Factores considerados

Como se ha comentado anteriormente, la red está formada por distintos tramos, y cada tramo lleva asociado un registro con sus características, tales como material, diámetro, etc.

Esta serie de características o factores, pueden ser de tres tipos:

Factores cuantitativos. Son valores que están definidos por valores numéricos medibles, tales como el diámetro, la longitud, el número de acometidas o la presión. La forma en la que estos factores se incluyen en el modelo de regresión es otorgando a la variable X_i el propio valor del factor.

Dentro de los factores cuantitativos se pueden encontrar los siguientes:

- Diámetro. Número que expresa el diámetro nominal del tramo en mm (milímetros).
- Año de Instalación. Valor numérico que expresa el año en el que el tramo entró en funcionamiento. Este factor entra en el ajuste a través de la Edad, que es el tiempo transcurrido desde el momento de la instalación de la tubería hasta el momento considerado.
- Longitud. Número que expresa la longitud real del tramo. Se representa en m (metros).
- Número de acometidas. Número que define la cantidad de intersecciones del tramo con otros objetos. Se define mediante la unidad “acometida”.
- Presión mínima. Define el valor menor de presión al que está sometido el tramo. Se mide en m.c.a. (metros de columna de agua).
- Presión media. Define el valor medio de presión al que está sometido el tramo. Se mide en m.c.a. (metros de columna de agua).
- Presión máxima. Define el valor máximo de presión al que puede estar sometido el tramo. Se mide en m.c.a. (metros de columna de agua).
- Número de incidencias previas acumuladas. Número que define la cantidad de incidencias sufridas por el tramo. Se mide en incidencias.
- NS: Nivel de servicio estimado por EMASESA. Número adimensional que define el nivel de riesgo asociado al tramo mediante el método desarrollado y utilizado por EMASESA.

Factores cualitativos. Son factores cuyos valores no se definen numéricamente. Dentro de los factores cualitativos se pueden encontrar los siguientes:

- **Tipo de red:** los tipos de red que se consideran dentro de la red de abastecimiento son:
 - Red Arterial y de Transporte de abastecimiento. Tipo de red perteneciente a la red de abastecimiento de EMASESA que tiene como función principal la conexión entre las distintas estaciones de tratamiento de aguas potables y los tramos principales de la red.
 - Red Secundaria de Abastecimiento. Constituye a todos aquellos tramos de menor capacidad que distribuyen el agua desde las conexiones de la RAA hasta los puntos de consumo, entre otras funcionalidades.
- **Material.** Cadena de caracteres que indica de qué material está fabricado el tramo. Los materiales que se registran en la hoja de datos extraída de las bases de datos de EMASESA para los tramos son:
 - Fibrocemento (FC). Material constituido por aglomerante reforzado con fibras orgánicas y/o sintéticas.
 - Hierro Fundido (HF). El Hierro Fundido o también llamado Fundición Gris es un tipo de aleación, formado generalmente por Carbono, Silicio, Manganeso, Fósforo y Azufre.
 - Fundición dúctil (FD). También conocida como Fundición Nodular, o Fundición de Grafito Esferoidal, es un tipo de fundición de hierro aleada con grafito.
 - Hormigón Armado con Chapa Metálica (HACH). Hormigón Armado Envuelto en una capa de Acero que le otorga comportamiento mixto y capacidad de impermeabilización si el acero es tratado frente a corrosión.

- Polietileno (PE). Polímero que se usa con gran frecuencia en sistemas de transporte de agua. Agrupa los tipos PE BD/MD.
 - PVC (PVC). Cloruro de Polivinilo, es el material derivado de la polimerización del monómero de cloruro de vinilo. Agrupa los tipos PVC.
 - Acero Inoxidable (AI). Aleación de acero con un mínimo de entre el 10% y 12% de cromo contenido en masa. Capacidades de impermeabilización y gran resistencia a la corrosión.
 - Otros. AH, CH, HG.
- **Ubicación:** los tipos de ubicación que se consideran para la red de abastecimiento son:
- En calzada. Tipo de ubicación que designa la posición del tramo en una zona sometida a carga de tráfico con gran frecuencia.
 - En terrizo. Tipo de ubicación que designa la posición del tramo en un terreno no urbanizado.
 - En acera Tipo de ubicación que designa la posición del tramo en cuestión en una zona peatonal sin cargas frecuentes de tráfico, pero con carga peatonal.
 - No hay datos. En las BBDD de EMASESA existe un número no despreciable de tramos que tienen vacío el campo “ubicación”, para estos tramos se ha creado esta ubicación ficticia.
- **Municipio.** Factor de localización que agrupa a cada tramo en un cierto núcleo poblacional. Los distintos municipios sobre los que se distribuye la red de abastecimiento de EMASESA son:
- Alcalá de Guadaíra
 - Alcalá del Río
 - Burguillos
 - Camas
 - Carmona
 - Coria del Río
 - Dos Hermanas
 - El Garrobo
 - El Ronquillo
 - Gelves
 - Guillena
 - La Puebla del Río
 - La Rinconada
 - Los Palacios y Villafranca
 - Mairena del Alcor
 - Palomares del Río
 - San Juan de Aznalfarache
 - Santiponce
 - Sevilla

- Tomares

➤ Distrito. Factor que designa la localización de un tramo en una determinada zona del municipio al que pertenece.

Dentro del municipio de Alcalá de Guadaíra se encuentran los distritos:

- AG_ALCALA DE GUADAIRA
- AG_LA GALBANA
- AG_LAS ENCINAS
- AG_PINOS DEL NEVERO
- AG_RANILLA-P.I. EL SOMBRERON
- AG_RUSTICA NORTE
- AG_SANTA GENOVEVA
- AG_SEVILLA GOLF
- AG_TORREQUINTO
- AG_ZONA INDUSTRIAL

Dentro del municipio de Alcalá del Río se encuentran los distritos:

- AR_ALCALA DEL RIO
- AR_EL VIAR
- AR_ESQUIVEL
- AR_ILIPA MAGNA
- AR_SAN IGNACIO DEL VIAR

Dentro del municipio de Burguillos se encuentra el distrito:

- BU_BURGUILLOS

Dentro del municipio de Camas se encuentran los distritos:

- CA_ZONA INDUSTRIAL
- CA_CAMAS
- CA_CARAMBOLO
- CA_OESTE
- CA_PANOLETA

Dentro del municipio de Carmona se encuentra el distrito:

- CAR_TORREPALMA

Dentro del municipio de Coria del Río se encuentra el distrito:

- CO_CORIA DEL RIO

Dentro del municipio de Dos Hermanas se encuentran los distritos:

- DH_ADRIANO
- DH_CONDEQUINTO
- DH_DOS HERMANAS

- DH_DOS HERMANAS CENTRO ESTE
- DH_DOS HERMANAS CENTRO OESTE
- DH_DOS HERMANAS ESTE
- DH_DOS HERMANAS NORTE
- DH_DOS HERMANAS OESTE
- DH_DOS HERMANAS SUR
- DH_ENTRENUCLEOS
- DH_FUENTE DEL REY
- DH_LA ISLA
- DH_MONTEQUINTO
- DH_PITAMO-HIPODROMO
- DH_PUNTA DEL VERDE

Dentro del municipio El Garrobo se encuentra el distrito:

- EG_EL GARROBO

Dentro del municipio de EL Ronquillo se encuentra el distrito:

- ER_EL RONQUILLO

Dentro del municipio de Gelves se encuentra el distrito:

- GE_GELVES

Dentro del municipio de La Puebla del río se encuentra el distrito:

- LP_LA PUEBLA DEL RIO

Dentro del municipio de La Rinconada se encuentran los distritos:

- RI_EL GORDILLO
- RI_ESPARTALES-AEROPOLIS
- RI_LA RINCONADA
- RI_NUEVA JARILLA
- RI_SAN JOSE DE LA RINCONADA
- RI_TARAZONA

Dentro del municipio de Mairena del Alcor se encuentran los distritos:

- MA_EL TORREON
- MA_MAIRENA DEL ALCOR

Dentro del municipio de Palomares del Río se encuentra el distrito:

- PA_PALOMARES DEL RIO

Dentro del municipio de San Juan de Aznalfarache se encuentra el distrito:

- SJ_SAN JUAN DE AZNALFARACHE

Dentro del municipio de Santiponce se encuentra el distrito:

- SA_SANTIPONCE

Dentro del municipio de Sevilla se encuentran los distritos:

- SE_BELLAVISTA-LA PALMERA
- SE_CASCO ANTIGUO
- SE_CERRO-AMATE
- SE_ESTE
- SE_LOS REMEDIOS
- SE_MACARENA
- SE_NERVION
- SE_NORTE
- SE_SAN PABLO-SANTA JUSTA
- SE_SUR
- SE_TRIANA
- SE_LOCAL (en este caso, se trata de un distrito auxiliar que recoge a los tramos que en principio no tienen asignado un distrito de los mencionados anteriormente).

Cabe destacar que los municipios Guillena, Los Palacios y Villafranca y Tomares, no cuentan con distritos propios.

Factores de identificación. Son una serie de factores que no tienen efecto en el proceso de resolución, pero que son muy importantes de cara a la identificación del tramo.

Los factores identificativos que se pueden encontrar son los siguientes:

- Código GIS: es un número identificador del tramo.
- ID sector. Es un número identificador del sector de fugas al que pertenece el tramo.
- Propiedad. Factor cuyo objetivo es destacar los tramos pertenecientes a EMASESA o que discurren por el terreno, pero son propiedad de empresas privadas.
- Sector de fugas. Define un conjunto de tramos que pueden ser aislados hidráulicamente cerrando determinadas válvulas.
- Indicador de borrado. Factor que toma valores de Sí o No, en función de si el tramo (que obviamente existió en algún momento) sigue existiendo o no.

Los factores de identificación no se incluyen dentro del proceso de regresión lineal, a diferencia de los cuantitativos y cualitativos, pero sí son importantes para identificar el tramo individualmente.

1.3.3 Conclusiones sobre el Uso de Datos desarrolladas por el GERM.

Estudios preliminares del GERM corroboraron que era mejor realizar la regresión por familias que por tramos, porque la calidad de los datos mejoraba.

El concepto de familia surge para intentar solventar los problemas derivados de la actualización de los códigos GIS de los tramos cuando la incidencia conlleva la sustitución de parte de un tramo original, que deviene en 3 tramos nuevos, de forma que sólo uno de ellos “hereda” el histórico de incidencias del tramo original, lo que penaliza la calidad de los datos.

Las conclusiones de los estudios preliminares que se realizaron son:

- Los resultados obtenidos con el análisis por familias son mejores que los resultados obtenidos por tramos. Esto se justifica porque al agrupar los tramos en familias se mejora la calidad de los datos.
- Es preferible hacer un ajuste independiente para las familias con incidencias previas y otro para

las familias sin incidencias previas. Ello es debido a que el factor más determinante en el ajuste es el número de incidencias previas, lo que hace que sólo se predigan incidencias en tuberías que ya han sufrido alguna. El realizar ajustes independientes incluye algunas familias sin incidencias previas en la predicción, mejorando los resultados.

La definición de familia se extrae del documento [2], y se define de la siguiente forma:

- Familia: *“es el conjunto de todos los tramos o registros que tienen en común una serie de propiedades. Se definen para dar fiabilidad a algunas características (longitud) y datos de los históricos. Estos tramos son generados automáticamente y se identifican con un código numérico que se genera de forma correlativa”*.

Para generar las familias, se agrupan los tramos en función de determinados factores que se clasifican según su tipología. Estos factores pueden ser:

- 1) **Factores que definen la familia:** son características o propiedades que todos los tramos que conforman la familia tienen en común. Estos son: Sector de Fuga, Material, Tipo de Red, Distrito, Año de Instalación, Diámetro, Ubicación, Indicador de Borrado.
- 2) **Factores que suman:** son factores cuantitativos, que caracterizan una familia mediante la suma de los valores de todos los tramos que la conforman. Dentro de este grupo están el factor Longitud, y el factor Número de Acometidas.
- 3) **Factores que se promedian:** son factores cuantitativos, que caracterizan una familia mediante el promedio de los valores de los tramos que la conforman. Dentro de este grupo están el factor Presión Media y el factor Fluctuación de Presión.

Es importante mencionar una particularidad para los factores relacionados con la presión: Presión Media y Fluctuación de Presión. En las hojas de datos provenientes de EMASESA se definen los factores Presión Mínima, Presión Media y Presión Máxima. Debido a la relación que existe entre estos factores y con la idea de reducir el número de factores, se decidió simplificarlos, manteniendo la presión media, y creando un nuevo factor Fluctuación de Presión como la diferencia para cada tramo entre la Presión Máxima y la Presión Mínima.

Teniendo en cuenta estas conclusiones, los análisis presentados en este TFG se realizarán exclusivamente sobre familias.

1.4 Objetivo.

Los objetivos fundamentales que se intentan conseguir en este proyecto son los siguiente:

A nivel académico:

- Conseguir la superación de los 12 créditos equivalentes al desarrollo del TFC (Trabajo Final de Carrera) o TFG (Trabajo Fin de Grado) del Plan de Estudios asignado al Grado en Ingeniería Civil, para la obtención del título de Graduado en Ingeniería Civil por la Universidad de Sevilla.

A nivel técnico, el objetivo principal es **realizar algunos análisis de la herramienta, para conocer las condiciones en las que debe utilizarse para que los resultados (predicciones) obtenidos tengan mayor fiabilidad.**

Para la consecución de dicho objetivo se han planteado los siguientes hitos:

- 1) Corroborar la posible relación entre factores cualitativos mediante un tratamiento estadístico basado en tablas de contingencia.
- 2) Analizar las características de las redes de abastecimiento de EMASESA para seleccionar el conjunto de tramos sobre los que realizar los estudios.
- 3) Analizar el efecto de la normalización de los factores (para aquellos susceptibles de tal operación). Es decir, transformarlos de forma que la distribución de cada factor resulte con media nula y desviación típica unitaria.

- 4) Corroborar la relación entre el coeficiente PF calculado con la regresión y la probabilidad de fallo real de la tubería.
- 5) Resumir las conclusiones de los estudios realizados en este TFG y proponer líneas de desarrollo futuras que pudieran contribuir a la mejora de las predicciones obtenidas.

1.5 Metodología.

En este apartado se explican algunos conceptos y herramientas que son importantes para conseguir una buena interpretación de los resultados. El TFG se estructura en 3 bloques independientes.

❖ Bloque de Análisis de Factores.

- *Generación de los datos de partida a partir de los datos extraídos de las Bases de Datos de EMASESA.*

Previamente a cualquiera de los hitos propuestos para alcanzar el objetivo técnico de este TFG, es necesario agrupar toda la información de las BBDD de EMASESA y generar **las familias**, para ello se utiliza el programa “Genera Datos”.

El programa permite elegir qué parámetros van a ser los representantes de los tramos a la hora de agruparlos en las diferentes familias. Siguiendo las recomendaciones sobre este aspecto que se relacionan en [2], estos parámetros o factores son: Sector de Fugas, Material, Tipo de Red, Distrito, Año de Instalación, Diámetro, Ubicación, e Indicador de Borrado.

- *Análisis de los Factores y Estudio en base al histórico de incidencias y distribución en la Red. Estudio de calidad de los registros. Estudio de Correlación de Factores.*

Para analizar la distribución de los factores en la red de abastecimiento de EMASESA, así como la influencia en el número de incidencias previstas se utiliza la macro “Análisis de Factores”, que permite obtener la información en forma de tablas o gráficas. Además, facilita la comprobación de la existencia de posibles correlaciones entre los factores.

- *Contraste de Resultados. Análisis de Bondad por quinquenios y Constatación de Previsión de Fallo frente a Probabilidad de Fallo.*

Para analizar el efecto de la normalización de los factores, así como para evidenciar la relación entre el PF y la probabilidad de fallo, se realizarán análisis con los subconjuntos de datos agrupados por quinquenios, utilizando los datos del año siguiente como comprobación. Los quinquenios comenzarán en 2003, primer año a partir del cual se registran las incidencias en el histórico, hasta 2017, manteniendo el año 2018 para verificación en el último quinquenio, ya que se trata del último año sobre el que se tiene constancia de incidencias reales en el histórico. Los resultados analizarán según las correspondientes matrices de confusión.

1.6 Contenido del trabajo.

El proyecto está dividido en cuatro capítulos.

El primer capítulo consiste en la introducción, donde se describe el marco técnico en el que se desarrolla el Trabajo Fin de Grado, los objetivos, y la metodología. Además, se incluye un apartado donde se describen brevemente en qué consisten los programas utilizados.

El segundo capítulo presenta un estudio exhaustivo de los datos que serán objeto de uso. En este capítulo se analizan cada uno de los factores que caracterizan los tramos que conforman la red de abastecimiento. Se realiza un análisis para comprobar si existen correlaciones entre factores cualitativos y, además, se realiza un procedimiento de filtrado, mediante el cual, se acotarán los rangos que pueden alcanzar determinados parámetros o factores, con el fin de eliminar ciertos valores que

puedan provocar una disminución de la calidad de los datos.

El tercer capítulo consiste en realizar una serie de tratamientos estadístico sobre los factores que definen los datos de partida con la finalidad de mejorar la calidad los resultados obtenidos, y poder obtener conclusiones acerca de la relación entre ellos.

Además, en dicho capítulo de corrobora la relación entre el coeficiente PF (Previsión de Fallo) obtenido como resultado del modelo de regresión lineal multiparámetro, y la Probabilidad de Fallo real.

Finalmente, en el último capítulo, correspondiente al capítulo 4, se dedicará a resumir brevemente el TFG, constatar que se han alcanzado los objetivos propuestos y se indicaran posibles desarrollos futuros.

1.7 Descripción de software utilizado

Este capítulo se dedica a describir los programas utilizados para el desarrollo del TFG. Se comentarán todas las funciones disponibles en los distintos Softwares empleados, la estructura de los resultados que se obtendrán, y algunas consideraciones básicas para poder ejecutar de forma correcta dichos programas.

La información acerca del uso de los distintos Softwares se ha extraído del documento [3].

Los tres programas cuyo uso se va a describir son: “**Software para Generación de Datos**”, “**Software para Análisis de Datos**”, y “**Software de Ajuste y Predicción**”.

1.7.1 Software para Generación de Datos

En la figura 6 se muestra la interfaz que presenta el Software para Generación de Datos.

El software para generación de datos tiene varias funciones:

- La primera función consiste en generar un fichero en formato Excel, agrupando la información de los dos ficheros extraídos de las BBDD de EMASESA, el relativo a las características recogidas en los registros que definen los tramos que componen la red, y el fichero que contiene el histórico de incidencias ocurridas en los tramos a partir de 2003, año en que comenzaron a registrarse dichas incidencias, hasta la actualidad; con la finalidad de transferir la información conjunta a los programas restantes para su posterior tratamiento.

Figura 6. Interfaz software Generación de Datos.

Además, también realiza la función de agrupar los tramos por familias, en base a ciertos factores que estos tengan en común y que están dispuestos a elección propia por parte del usuario (predefinidos en subapartados 1.3.2). Para la generación de familias, se enumeran los pasos a seguir:

1. En primer lugar, es necesario haber generado previamente el fichero de tramos.

2. Una vez generado el fichero de tramos, se definen qué factores que definen la familia.
3. Tras definir los factores, se ejecuta mediante el botón “Genera FAMILIAS”. En el programa “Genera Datos” resultan dos hojas:
 - a) Una primera hoja con el nombre de “Informe de familias_Abto”, que contiene información acerca de la cantidad de familias que se han generado, del tiempo de procesamiento, y del tiempo empleado, número de tramos, número de familias y número de iteraciones que han tenido lugar durante el procesamiento clasificados en función del Sector de Fugas.
 - b) Una segunda hoja donde se muestran todas las variables de los tramos de abastecimiento, llamada “Var_Abto”.

Por otro lado, genera automáticamente en los dos programas restantes el fichero de información perteneciente a las familias, habilitando la información para su tratamiento.

Para generar el fichero de información acerca de los tramos, incluyendo la información sobre estos y su histórico de incidencias, no es necesario realizar ningún ajuste en el software. Simplemente se define el año de comienzo de históricos, y el identificador de cada tramo (“Tramo GIS”). Por último, se detallan los nombres de las dos hojas que contienen la información extraída de las bases de datos de EMASESA (para que el programa se dirija a esas hojas), y se ejecuta mediante el botón “Genera Datos”, de color azul que aparece en la parte central derecha de la imagen.

El resultado es una hoja llamada “Datos Abastecimiento”, cuya pestaña es de color verde y se encuentra en la zona inferior de la ilustración.

Además, se introduce el fichero con la información acerca de los tramos en los dos programas restantes disponiéndola para su tratamiento.

Por otro lado, como información adicional, aparece en la parte inferior el tiempo empleado en generar tanto la hoja de tramos, como la de familias.

En el último proceso que se ejecutó, el tiempo empleado fue 564 segundos en el caso de los tramos y de 3281 segundos en el caso de las familias, equivalente a 1 hora y 4 minutos. En este dato se puede comprender la magnitud de los datos con los que se trabaja.

1.7.2 Software para Análisis de Factores

El software “Análisis de Factores” consta de 5 hojas principales. Cada una de ellas presenta una serie de funciones que a continuación se describen:

- La primera hoja, denominada “Análisis de la Red” permite llevar a cabo un análisis global de los datos para un filtro determinado. Por ejemplo, en la figura 7 se analizarían los datos para los factores cualitativos “Material” y “Distrito”.

 		Análisis de la red de Abastecimiento		
Fila comienzo variables	6	TIPO DE RED GLOBAL	Abastecimiento	
Localización Factores cualitativos		'Var_Abto'!\$A\$12:\$A\$25	Fila	Colum.
Factor cualitativo 1	2	Material	4	9
Factor cualitativo 2	10	Distrito	4	3
Variable cuantitativa a analizar	5	Longitud		
Factor de control de borrado	14	Indicador de Borrado		
Fila comienzo registros ; Columna comienzo históricos	3	21		
Nº total de filas		23412		
Hoja de datos	F	D_Fa_Abto	'D_Fa_Abto'!\$A\$2:\$T\$2	
Hoja de resultados		Material VS Distrito		
Hoja de datos de Variables		Var_Abto		
Tiempo de ejecución (s)		48	EJECUTAR	

Figura 7. Interfaz software Análisis de factores. Análisis de la Red.

- En segundo lugar, la hoja “Análisis de incidencias”, permite realizar un análisis de la evolución de las incidencias acumuladas permite realizar un análisis de la evolución de las incidencias acumuladas en un período concreto (definido por el usuario) del conjunto de tramos que cumplen determinadas condiciones (material, zona, etc. definidas por el usuario). Se puede apreciar en la figura 8, la estructura relativa a esta hoja.

Análisis de las incidencias acumuladas por zonas o por los distritos de una zona									
Fila comienzo variables	6	TIPO DE RED GLOBAL			Abastecimiento				
Localización Factores cualitativos		'Var_Abto'ISA512:SA525			Fila	Column.	Num.	Localización	
MATERIAL	9	2	TODOS	TODOS	4	9	9	'Var_Abto'ISK55:SK513	
Zona	6	10	Dos Hermanas	DH	4	3	75	'Var_Abto'ISF55:SF579	
Tipo de red		8	Red secundaria de abastecimiento		4	13	3	'Var_Abto'ISN55:SN57	
Longitud		5	Longitud						
Año instalación		4	Año Instalación						
Fila comienzo reg. ; Nº total filas		3	23411						
EJECUTAR									
Incidencias acumuladas para el periodo 2010-2018		29	2010						
Año de comienzo de Historicos		37	2018						
Año final de historico		21	2002						
		39	2020						
Hoja de datos para tramos o familias (T ó F)		F	D_Fa_Abto		'D_Fa_Abto'ISA52:ST52				
Hoja para grafico			Fig 11						
Hoja de datos de Variables			Var_Abto						
Tiempo de ejecución (s)			108						
Nota: Se utilizan todos los registros del fichero de datos incluido los que tienen SI en el indicador de borrado									
Nota: El primer año del periodo es el de referencia (incidencias cero)									

Figura 8. Interfaz software Análisis de Factores. Análisis de Incidencias.

- En tercer lugar, la hoja denominada “Efecto de las Incidencias Previas” permite conocer las incidencias acumuladas y por longitud instalada para un año concreto para un periodo definido distinguiéndose si ha habido o no incidencias previas. La composición de la hoja relativa a “Efecto de las Incidencias Previas” se muestra en la figura 9.

Análisis de las incidencias acumuladas por zonas o por los distritos de una zona									
Fila comienzo variables	6	TIPO DE RED GLOBAL			Abastecimiento				
Localización Factores cualitativos		'Var_Abto'ISA512:SA525			Fila	Column.	Num.	Localización	
MATERIAL	9	2	TODOS	TODOS	4	9	9	'Var_Abto'ISK55:SK513	
Zona	6	10	Dos Hermanas	DH	4	3	75	'Var_Abto'ISF55:SF579	
Tipo de red		8	Red secundaria de abastecimiento		4	13	3	'Var_Abto'ISN55:SN57	
Longitud		5	Longitud						
Año instalación		4	Año Instalación						
Fila comienzo reg. ; Nº total filas		3	23411						
EJECUTAR									
Incidencias acumuladas para el periodo 2010-2018		29	2010						
Año de comienzo de Historicos		37	2018						
Año final de historico		21	2002						
		39	2020						
Hoja de datos para tramos o familias (T ó F)		F	D_Fa_Abto		'D_Fa_Abto'ISA52:ST52				
Hoja para grafico			Fig 11						
Hoja de datos de Variables			Var_Abto						
Tiempo de ejecución (s)			108						
Nota: Se utilizan todos los registros del fichero de datos incluido los que tienen SI en el indicador de borrado									
Nota: El primer año del periodo es el de referencia (incidencias cero)									

Figura 9. Interfaz software Análisis de Factores. Efecto de las Incidencias Previas.

- En cuarto lugar, la hoja “1Factor”, tiene la función de realizar análisis por factores individuales, es decir, permite visualizar la evolución de las incidencias acumuladas y por unidad de longitud para un factor concreto.

Se puede apreciar la estructura de la hoja “1Factor” en la figura 10.

Análisis de los factores sobre las incidencias acumuladas en el periodo 2007-2016										
Fila comienzo variables	E	TIPO DE RED GLOBAL			Abastecimiento					Variable cualitativa
Factor 1	2	Material	3						Material	9
Longitud	5				Fila	Column.	Num.	Localizacion	1	AI
MATERIAL	9	2	TODOS	TODOS	4	9	9	'Var_Abto'ISK55:SK513	2	PC
Zona	6	10	Sevilla	SE	4	6	19	'Var_Abto'ISF55:SF523	3	PD
Tipo de red		8	Red secundaria de abastecimiento		4	13	3	'Var_Abto'ISN55:SN57	4	HACH
									5	HF
									6	PE
Fila comienzo registros ; Columna comienzo históricos		3	23411						7	PKC
Nº total de filas		25	2006						8	VA
Incidencias acumuladas para el periodo 2006-2016		35	2016						9	TODOS
Año de comienzo de Historicos		21	2002							
Año final de historico		39	2020							
Hoja de datos		F	D_Fa_Abto		'D_Fa_Abto'ISA52:ST52					
Hoja para grafico			Material							
Hoja de datos de Variables			Var_Abto							
Tiempo de ejecución (s)			5							
Nota: El primer año del periodo es el de referencia (incidencias cero)										

Figura 10. Interfaz software de Análisis de Factores. 1Factor.

- En último lugar, la hoja denominada “2Factores”, permite realizar análisis por parejas de factores, desarrollando resultados que permiten observar posibles correlaciones entre factores. En la figura 11 se muestra la configuración de la hoja para el caso de realizar el análisis de incidencias entre los factores ubicación y diámetro, respectivamente.

Cabe destacar que todas las hojas descritas tienen capacidad para realizar el análisis sobre tramos o sobre familias.

Análisis comparativo entre los factores									
Fila comienzo variables	6	TIPO DE RED GLOBAL				Abastecimiento			
Factor 1	9	Ubicación	6	4	19	5			
Factor 2	3	Diámetro							
MATERIAL	3	2	Fundición Dúctil	FD	4	9	9	Var_Abto\1SK55:SK513	
Zona	6	10	Sevilla	SE	4	6	19	Var_Abto\1SF55:1SF523	
Tipo de red	8		Red secundaria de abastecimiento		4	13	3	Var_Abto\1SN55:1SN57	
Fila comienzo registros ; Nº total de filas	3	23411							
Incidencias acumuladas para el periodo 2006-2016	25	2006							
Año de comienzo de Historicos	35	2016							
Año final de historico	21	2002							
Año final de historico	39	2020							
Hoja de datos	F	D_Fa_Abto		D_Fa_Abto\1SA52:1ST52					
Hoja para grafico		Ubic vs Diám							
Hoja de datos de Variables		Var_Abto							
Tiempo de ejecución (s)		3							

Figura 11. Software de Análisis de Factores. 2Factores.

1.7.3 Software de Ajuste y Predicción

El Software de Ajuste y Predicción tiene como objetivo generar y resolver el sistema de regresión multiparámetro por mínimos cuadrados, además de calcular los resultados mediante la herramienta “Matriz de confusión” para el análisis de bondad de las predicciones.

Por otro lado, dicho Software también extrae los coeficientes de ajuste para el estudio de peso de los factores sobre la probabilidad de ocurrencia de una incidencia en familias y tramos.

La interfaz que presenta este programa se muestra en la figura 12.

ESTIMACIÓN PRELIMINAR DEL RIESGO DE INCIDENCIAS DE UNA RED DE TUBERÍAS EN BASE A HISTÓRICOS DE FUNCIONAMIENTO									
Tipo de Red Global a analizar	Abastecimiento				EJECUTAR				
Nombre hojas de datos	D_Tr_Abto	D_Fa_Abto	D_In_Abto	Var_Abto					
Fila en la que comienzan los registros	3	3	3	4					
Nº de registros totales	114380	23411	114380						
Nº de factores considerados fichero datos- Maximo	17	26							
Año de referencia de historicos de rotura	2002	21							
Año comienzo-final registros de roturas	2003	2020							
Fila para comienzo definición variables y Nº de variables iniciales	35	15							
Hay Resultados de Inspecciones	NO								
Definición de parámetros									
Fase de ajuste (año 0 y año N _{ajuste})	2003	2018	OK						
Fase de verificación o predicción (año 0 y Año N _{verif})	2018	2019	Verificación	Fila	Column.	Num.	Localizacion		
Selección del Material	TODOS	TODOS		4	9	9	Var_Abto\1SK55:SK513		
Selección Zona	SE	Sevilla		4	6	19	Var_Abto\1SF55:1SF523		
Selección tipo de red	RSA	Red secundaria de abastecimiento		4	13	3	Var_Abto\1SN55:1SN57		
Nivel de confianza para las predicciones (>90%)	97.5%								
Longitud mínima de Supertramo-Tramos GIS para predicción (m)	1.5	1.5							
Resultados de las Inspecciones									
Tramos	Familia	Global							
Código	Concepto	To							
1	Sustitución prioritaria	3							
2	Sustitución recomendada	5							
3	Sustitución a programar	10							
4	Buen Estado	0							
0	Sin inspección	0							
To: Periodo de año para el que se espera el fallo									

Figura 12. Interfaz del Software de Ajuste y Predicción.

En la figura 13 se presenta la estructura del programa donde se definen los factores que se quieren incluir en el modelo de regresión.

Es necesario indicar la tipología del factor que se va a incluir (definir si se trata de un factor cuantitativo o cualitativo).

Los factores que se han definido para el desarrollo de los dos capítulos siguientes son prácticamente los mismos, entre los que se encuentran:

- Diámetro (Cuantitativo)
- Material (Cualitativo)

Los resultados que se obtienen tras ejecutar el programa se organizan en las siguientes hojas:

- a) Hoja “R_Tr_Abto”: en esta hoja se recogen los resultados referentes a la previsión de fallo (PF) asignada a los tramos tras ejecutar el proceso de ajuste y predicción. Además, se recoge también una tabla con el valor y porcentaje de longitud de tramos asociada a cada nivel de riesgo, y el número de tramos en esta misma clasificación.
- b) Hoja “R_Fa_Abto”: Ídem que a) referida a las familias, en lugar de a los tramos.
- c) Hoja “Verif_Abto”: en esta hoja aparecen los resultados asociados al análisis de bondad, representando los valores obtenidos mediante la “Matriz de Confusión”, para familias y tramos.
- d) Hoja “R_GL_Abto”: aparecen en esta hoja los resultados asociados al ajuste del sistema de ecuaciones. En este grupo de resultados se encuentran la pendiente, término independiente, calidad del ajuste R^2 , etc.
- e) Hoja “Coef_Ajuste”: en esta hoja se muestran los coeficientes de ajuste.

2 ESTUDIO DE FACTORES

En este capítulo se desarrolla el estudio de los factores, para lo cual se analiza la distribución de la longitud de red en función de estos, para la red de abastecimiento de EMASESA, así como las posibles interrelaciones existentes entre los factores y con el histórico de incidencias disponible.

La conclusión se centra en una vez desarrollado este análisis, aplicar un tratamiento de filtrado de *outliers* de los datos, con el fin de mejorar la calidad de los mismos. Para Cuantificar la mejora de calidad de los datos se realizará un análisis de bondad y se compararán los parámetros obtenidos mediante la matriz de confusión definidos en el apartado 1.3.1 Solución Implementada.

El capítulo se estructura en dos apartados:

- El primer apartado 2.1 Análisis de Factores Cualitativos, consiste en analizar los factores cualitativos, con el fin de encontrar mediante **Tablas de Contingencia** posibles correlaciones entre los factores, que podrían introducir inestabilidades en el análisis de regresión.
- El segundo apartado, 2.2 Efecto de los factores sobre la aparición de incidencias, se centra en realizar un análisis de la distribución de los factores, así como de la relación existente entre el histórico de incidencias y los propios factores.
- El tercer apartado, 2.3 Estudio del efecto de filtrado de *outliers*, incluye el análisis de bondad comparando las predicciones con los datos que han sido seleccionados mediante el proceso de filtrado, y las obtenidas con los datos originales.

Para realizar esta comparación se ha decidido tomar como fase de ajuste periodos designados por quinquenios, y como período de predicción/comprobación el año siguiente al quinquenio en cuestión, es decir, el primer quinquenio comprenderá desde el año 2003 (primer año del que se dispone de registros de incidencias) hasta el año 2008, y posteriormente se predecirá y comprobará para el año 2009. La fase de ajuste correspondiente al segundo quinquenio abarcará desde el año 2004, hasta el año 2009, y posteriormente se predecirá y comparará para el año 2010, y así consecutivamente hasta que el último año para predecir sea el 2018 (último año del que se tiene constancia de incidencias en los datos disponibles), cuyo quinquenio para fase de ajuste sería el que abarca desde el año 2012 hasta el año 2017.

El motivo por el cual se realiza el cálculo de los parámetros de esta manera es debido a que permiten obtener muchas comprobaciones para comparar la calidad de los datos (filtrados y originales) utilizando el mismo histórico. Además, permite analizar la evolución con los años de la calidad de los datos de ambos conjuntos. El estudio podría haberse realizado por décadas (o cualquier otro número de años).

2.1 Análisis de Factores Cualitativos

El objetivo sobre el que se fundamenta este apartado se basa en hallar posibles grados de asociación o correlaciones entre los factores cualitativos que definen los registros de datos obtenidos a partir de las bases de datos de EMASESA. Este procedimiento también se podría aplicar a los valores cuantitativos sin más que definir categorías (por ejemplo: diámetro: $0 < D \leq 50$, $50 < D \leq 100$, etc.).

Este propósito se plantea debido a que el hecho de utilizar factores relacionados entre sí genera una inestabilidad en los cálculos.

En este apartado se pretende desarrollar un proceso de análisis estadístico mediante la herramienta **Tablas de Contingencia** [4], con la idea de estudiar posibles relaciones entre las variables o factores cualitativos que definen la estructura de datos con la que se trabaja.

2.1.1 Tablas de Contingencia.

Se denominan tablas de contingencia a tablas de doble entrada conformadas por la información obtenida a través de la observación de dos caracteres nominales. Estas observaciones son denominadas distribuciones de frecuencias. [4]

En la figura 16 se muestra la estructura de la que dispone una tabla de contingencia.

		Atributo B		Modalidad				Total
		Atributo A		B_1	B_2	...	B_k	
Modalidad	A_1	n_{11}	n_{12}	...	n_{1k}	$n_{1\bullet}$		
	A_2	n_{21}	n_{22}	...	n_{2k}	$n_{2\bullet}$		
		
	A_h	n_{h1}	n_{h2}	...	n_{hk}	$n_{h\bullet}$		
Total		$n_{\bullet 1}$	$n_{\bullet 2}$...	$n_{\bullet k}$	N		

Figura 16. Estructura Tabla de contingencia

Los significados de las distintas variables que aparecen en la tabla son los siguientes:

- **Atributos.** Los atributos definen los distintos factores cualitativos que van a ser objeto de estudio. Entre estos factores podemos encontrar Ubicación, Distrito, Material, etc.
- **Modalidades.** Las modalidades que puede adoptar cada atributo representan los distintos valores que pueden adoptar cada uno de los factores cualitativos que definen el registro de información. Por ejemplo, el factor Ubicación puede adoptar los valores de En Acera, En Calzada, En Terrizo, No hay datos, respectivamente.
- **Valores n_{ij} .** Los distintos valores n_{ij} que están definidos por las distintas modalidades que definen los atributos, son en este caso particular el número de incidencias asociadas a la modalidad i del atributo A, y a la modalidad j , del atributo B, en cuestión.

Así, por ejemplo, en el caso de que en la tabla estuvieran representados los atributos o factores A, Ubicación y B, Material, estando en la posición A_1 el valor de En Acera, y en B_1 , el valor de FC, n_{11} se correspondería con el número de incidencias registradas en el periodo en cuestión atribuidas a los tramos que están constituidos por FC, y se encuentran en la ubicación de En Acera.

Al total de incidencias registradas en la tabla se le denomina mediante el parámetro N .

Los valores de $n_{i\bullet}$ y $n_{\bullet j}$ representan la suma de incidencias para una modalidad concreta de un atributo. A este valor se le conoce como "Frecuencias marginales".

Existen 2 tipos de tablas de contingencia, las tablas de 2×2 y las tablas de $h \times k$.

Las tablas de contingencia de 2×2 son tablas donde los dos atributos cuya información entra en la tabla solo disponen de dos modalidades cada uno. Como en el caso de los factores cualitativos con los que se cuentan en los datos disponibles para este estudio pueden tener más de dos modalidades, se acudirá al uso de las tablas de contingencia $h \times k$, donde el primer atributo puede tomar h modalidades y el segundo atributo puede tomar k modalidades.

Los datos de entrada en la tabla, como se ha comentado anteriormente, son las incidencias registradas para ambas modalidades, que se han obtenido mediante la macro "Análisis de Datos", concretamente en la hoja de "2factores", donde se facilita el número de incidencias registradas en el histórico en función de las modalidades de dos factores concretos, ya sean cuantitativos o cualitativos.

Para comprobar si dos factores están interrelacionados o no, se implementarán dos métodos estadísticos independientes basados en dos coeficientes: el primer método se corresponde con el coeficiente de contingencia χ^2 , y el segundo método se corresponde con el coeficiente de contingencia K. Pearson.

- **Coefficiente de Contingencia χ^2 .** Se trata de una prueba estadística que permite identificar la asociación entre dos variables categóricas. Este coeficiente se utiliza para contrastar la hipótesis de independencia.

Según la metodología de dicho procedimiento, en primer lugar, es necesario establecer la **hipótesis teórica** (véase [4]), o de partida, la cual considera que las dos variables de estudio son independientes.

Por otro lado, se considera también la hipótesis opuesta o **hipótesis alternativa** (véase [4]), la cual considera que las dos variables de estudio sí están relacionadas.

En segundo lugar, es necesario obtener las frecuencias marginales tanto de fila como de columna de la tabla de contingencia.

Una vez calculados dichos valores, se genera otra tabla de la misma estructura que la primera, donde se calcularán las frecuencias esperadas mediante la siguiente expresión:

$$E_{ij} = \frac{n_{.i} * n_{.j}}{N}$$

Siendo $n_{.i}$ y $n_{.j}$ las frecuencias marginales calculadas en la tabla anterior para cada modalidad i del atributo A y cada modalidad j del atributo B.

N constituye el total de incidencias para ambos atributos o factores.

Una vez calculadas dichas frecuencias teóricas se calcula el coeficiente de contingencia χ^2 , también denominado cuadrado de la contingencia. La expresión por la que se define este coeficiente es la siguiente:

$$\chi^2 = \sum_{i=1}^h \sum_{j=1}^k \frac{(n_{ij} - E_{ij})^2}{E_{ij}}$$

Este coeficiente es siempre positivo y si las variables fueran independientes tomaría el valor teórico de 0.

En este sentido se tiene en cuenta que el estadístico χ^2 , sigue una distribución Ji cuadrado con $(h - 1) * (k - 1)$ grados de libertad: $\chi^2_{(h-1)*(k-1)}$

Se aceptará independencia de las variables si el $p - valor$ es mayor que el nivel de significación (α).

$$p - valor \rightarrow p = P\{X > \chi^2, \text{ siendo } X \sim \chi^2_{(h-1)(k-1)}\}$$

O de forma equivalente, si se verifica:

$$\chi^2 < \chi^2_{(h-1)(k-1), \alpha}$$

Siendo α el nivel de significación.

- **Coefficiente de contingencia K. Pearson.** Una vez contrastada la hipótesis de independencia, en el caso de que los resultados muestren que existe relación entre los factores, es necesario representar dicha relación en una escala cuantitativa.

Debido a que los límites del coeficiente de contingencia χ^2 varían dependiendo de los datos no es muy adecuado para constituir por sí mismo un coeficiente. Para ello se utiliza el coeficiente de contingencia K. Pearson, que se calcula de la siguiente forma:

$$C = \sqrt{\frac{\chi^2}{N + \chi^2}}$$

Este coeficiente varía entre 0 y 1, de modo que:

- Si los atributos son independientes el valor de C es 0.

- Cuanto mayor sea el grado de asociación más se acerca a 1.

En realidad, $C = 1$ no se alcanza más que en casos teóricos de infinitas modalidades, pero con este coeficiente se puede cuantificar el grado de relación entre dos atributos o factores.

2.1.2 Desarrollo del Proceso y Resultados.

El desarrollo de las tablas de contingencia para establecer las posibles relaciones entre factores cualitativos se va a realizar conforme al histórico de incidencias registrado en el periodo que abarca desde el año 2005 hasta el año 2018.

En primer lugar, como se ha mencionado en el apartado anterior, se contrastará la hipótesis de independencia de los dos factores entre sí, y en caso de que se cumpla la hipótesis alternativa, se llevará a cabo la cuantificación con el coeficiente K. Pearson de la relación entre ellos.

Los factores que se van a estudiar son los factores cualitativos que intervienen en el análisis.

Entre ellos se encuentran:

- **Material**
- **Ubicación**
- **Distrito**

No se incluyen los factores cualitativos “Municipio” y “Tipo de Red”, debido a que las incidencias se han obtenido concretamente para la Red Secundaria de Abastecimiento y el Municipio de Sevilla.

○ Relación entre los factores “Material” y “Ubicación”

En las tablas 4, 5 y 6 se muestran las frecuencias observadas, frecuencias esperadas, y términos para el sumatorio con el objetivo de obtener el valor de χ^2 , para el caso de la relación entre los factores Material y Ubicación.

FRECUENCIAS OBSERVADAS					
MATERIAL	UBICACIÓN				TOTAL
	En acera	En calzada	En terrizo	No hay datos	
FC	637	257	56	1470	2420
FD	1661	903	78	3458	6100
HACH	3	4	4	15	26
HF	462	237	20	1137	1856
PE	123	52	11	465	651
PVC	4	1	0	16	21
VA	2	4	1	12	19
TOTAL	2892	1458	170	6573	11093

Tabla 4. Frecuencias Observadas. Factores Material y Ubicación.

FRECUENCIAS ESPERADAS					
MATERIAL	UBICACIÓN				TOTAL
	En acera	En calzada	En terrizo	No hay datos	
FC	630.9059767	318.0708555	37.08645091	1433.936717	2420
FD	1590.300189	801.7488506	93.48237627	3614.468584	6100
HACH	6.778328676	3.417290183	0.398449473	15.40593167	26
HF	483.8683855	243.9419454	28.44316235	1099.746507	1856
PE	169.7189218	85.56368881	9.976561796	385.7408275	651
PVC	5.47480393	2.760118994	0.321824574	12.4432525	21
VA	4.953394032	2.497250518	0.291174615	11.25818083	19
TOTAL	2892	1458	170	6573	11093

Tabla 5. Frecuencias Esperadas. Factores Material y Ubicación.

TÉRMINOS SUMATORIO COEFICIENTE DE CONTINGENCIA CHI CUADRADO

MATERIAL	UBICACIÓN			
	En acera	En calzada	En terrizo	No hay datos
FC	0.05886316	11.7258445	9.645634192	0.906985906
FD	3.143094156	12.78679133	2.564162194	6.773448751
HACH	2.106089608	0.099362569	32.55410558	0.010695914
HF	0.988339593	0.197549484	2.506296229	1.261947864
PE	12.86042614	13.16587939	0.10498865	16.28558858
PVC	0.397283019	1.122422214	0.321824574	1.016651616
VA	1.760921149	0.904296941	1.72553994	0.048879626

Tabla 6. Términos Sumatorio Coeficiente de Contingencia Chi cuadrado. Factores Material y Ubicación.

$$\chi^2 = 137.0439$$

$$n \text{ (grados de libertad)} = (7 - 1) \times (4 - 1) = 18$$

$$\text{Nivel de riesgo o de significación (5\%)} = 0.05$$

$$\chi^2_{7,0.05} = 28.87 \quad \text{¿} \chi^2 > \chi^2_{7,0.05} \text{?} \rightarrow \text{Sí}$$

Debido a que el coeficiente χ^2 es mayor que el valor teórico, se concluye con la existencia de una relación entre estos dos factores.

Finalmente se calcula el coeficiente de K. Pearson y se cuantifica la relación.

$$C = \sqrt{\frac{\chi^2}{N + \chi^2}} = \sqrt{\frac{137.0439}{137.0439 + 11093}} = 0.11$$

El coeficiente de contingencia χ^2 , muestra que hay dependencia, pero tras analizar el resultado del coeficiente K. Pearson, se comprueba que esta dependencia es muy baja.

Las tablas para la distribución de χ^2 , para la obtención del coeficiente teórico, se adjunta al final de este documento.

○ **Relación entre los factores “Material” y “Distrito”**

En las tablas 7, 8 y 9, se muestran las frecuencias observadas, frecuencias esperadas, y términos para el sumatorio con el objetivo de obtener el valor de χ^2 , para el caso de la relación entre los factores Material y Distrito.

FRECUENCIAS OBSERVADAS

DISTRITO	MATERIAL							TOTAL
	FC	FD	HACH	HF	PE	PVC	VA	
SE_BELLAVISTA-LA PALMERA	167	440	3	84	86	1	1	782
SE_CASCO ANTIGUO	9	970	0	294	83	2	4	1362
SE_CERRO-AMATE	287	617	3	100	52	0	0	1059
SE_ESTE	426	476	2	37	92	2	0	1035
SE_LOCAL	145	181	5	134	29	0	2	496
SE_LOS REMEDIOS	124	255	6	95	36	4	4	524
SE_MACARENA	291	479	1	218	32	0	2	1023
SE_NERVION	71	644	0	323	23	0	0	1061
SE_NORTE	378	473	5	64	96	0	1	1017
SE_SAN PABLO-SANTA JUSTA	252	498	0	159	25	1	0	935
SE_SUR	189	620	0	167	42	0	0	1018
SE_TRIANA	81	447	1	181	55	11	5	781
TOTAL	2420	6100	26	1856	651	21	19	11093

Tabla 7. Frecuencias Observadas. Factores Material y Distrito.

FRECUENCIAS ESPERADAS								
DISTRITO	MATERIAL							
	FC	FD	HACH	HF	PE	PVC	VA	TOTAL
SE_BELLAVISTA-LA PALMERA	170.597674	430.018931	1.83286757	130.838547	45.8921843	1.48039304	1.33940323	782
SE_CASCO ANTIGUO	297.127919	748.958803	3.19228342	227.879924	79.9298657	2.57838276	2.3328225	1362
SE_CERRO-AMATE	231.026774	582.340215	2.48210583	177.18417	62.1481114	2.00477779	1.81384657	1059
SE_ESTE	225.791039	569.142703	2.42585414	173.168665	60.7396556	1.95934373	1.77273957	1035
SE_LOCAL	108.205174	272.74858	1.16253493	82.987109	29.1080862	0.93897052	0.84954476	496
SE_LOS REMEDIOS	114.313531	288.145677	1.2281619	87.6718651	30.7512846	0.99197692	0.89750293	524
SE_MACARENA	223.173172	562.543947	2.3977283	171.160912	60.0354277	1.9366267	1.75218606	1023
SE_NERVION	231.463085	583.440007	2.48679347	177.518796	62.2654827	2.00856396	1.81727215	1061
SE_NORTE	221.864239	559.244569	2.38366537	170.157036	59.6833138	1.92526819	1.74190931	1017
SE_SAN PABLO-SANTA JUSTA	203.97548	514.15307	2.1914721	156.437393	54.8710899	1.77003516	1.60146038	935
SE_SUR	222.082394	559.794465	2.38600919	170.324349	59.7419995	1.92716127	1.7436221	1018
SE_TRIANA	170.379519	429.469035	1.83052375	130.671234	45.8334986	1.47849995	1.33769044	781
TOTAL	2420	6100	26	1856	651	21	19	11093

Tabla 8. Frecuencias Esperadas. Factores Material y Distrito.

TÉRMINOS SUMATORIO COEFICIENTE DE CONTINGENCIA CHI CUADRADO							
DISTRITO	MATERIAL						
	FC	FD	HACH	HF	PE	PVC	VA
SE_BELLAVISTA-LA PALMERA	0.07587008	0.23166827	0.74320596	16.767608	35.0525239	0.15588933	0.08600438
SE_CASCO ANTIGUO	279.400528	65.2361794	3.19228342	19.1849476	0.11792494	0.12974281	1.19146691
SE_CERRO-AMATE	13.5612077	2.06288471	0.1080592	33.6226206	1.65707635	2.00477779	1.81384657
SE_ESTE	177.525326	15.2432123	0.0747579	107.074252	16.0884865	0.00084362	1.77273957
SE_LOCAL	12.5119635	30.8628626	12.6672651	31.3580637	0.00040135	0.93897052	1.55794883
SE_LOS REMEDIOS	0.82079243	3.81277951	18.5402582	0.61252901	0.89586545	9.12138441	10.7247428
SE_MACARENA	20.6139408	12.4071925	0.81478973	12.8177638	13.0920231	1.9366267	0.03504864
SE_NERVION	111.241936	6.28601515	2.48679347	119.225577	24.7613616	2.00856396	1.81727215
SE_NORTE	109.8797	13.3003091	2.87171469	66.2289174	22.0983322	1.92526819	0.31599201
SE_SAN PABLO-SANTA JUSTA	11.3070185	0.50747855	2.1914721	0.04197817	16.2614231	0.33499569	1.60146038
SE_SUR	4.92810255	6.47506661	2.38600919	0.06488382	5.26896568	1.92716127	1.7436221
SE_TRIANA	46.8876683	0.71561562	0.37681549	19.3844092	1.83326062	61.3182049	10.0266183

Tabla 9. Términos Sumatorio Coeficiente de Contingencia Chi cuadrado. Factores Material y Distrito.

$$\chi^2 = 1670.349$$

$$n(\text{grados de libertad}) = (12 - 1)(7 - 1) = 66$$

$$\text{Nivel de riesgo o de significación (5\%)} = 0.05$$

$$\chi^2_{7,0.05} = 85.963 \quad \text{¿}\chi^2 > \chi^2_{7,0.05}\text{?} \rightarrow \text{Sí}$$

Debido a que el coeficiente χ^2 es mayor que el valor teórico, se concluye con la existencia de una relación entre estos dos factores.

Finalmente se calcula el coeficiente de K. Pearson y se cuantifica la relación.

$$C = \sqrt{\frac{\chi^2}{N + \chi^2}} = \sqrt{\frac{1670.349}{1670.349 + 11093}} = 0.3618$$

Luego, existe relación entre dichos factores, es decir, no son independientes, y esta relación ya es considerable, de en torno a 0.4. Se puede afirmar que el factor material y el factor distrito están relacionados entre sí, aunque esta correlación no es muy fuerte.

○ **Relación entre los factores “Ubicación” y “Distrito”**

En las tablas 10, 11 y 12 se muestran los resultados de las frecuencias observadas, teóricas y los coeficientes del sumatorio de χ^2 para el estudio de relación de los factores Ubicación y Distrito.

FRECUENCIAS OBSERVADAS					
DISTRITO	UBICACIÓN				TOTAL
	En acera	En calzada	En terrizo	No hay datos	
SE_BELLAVISTA-LA PALMERA	184	71	24	503	782
SE_CASCO ANTIGUO	250	329	8	775	1362
SE_CERRO-AMATE	255	140	9	655	1059
SE_ESTE	290	109	22	614	1035
SE_LOCAL	90	33	5	368	496
SE_LOS REMEDIOS	162	61	24	277	524
SE_MACARENA	317	129	9	568	1023
SE_NERVION	329	121	10	601	1061
SE_NORTE	240	131	20	626	1017
SE_SAN PABLO-SANTA JUSTA	308	118	11	498	935
SE_SUR	289	121	19	589	1018
SE_TRIANA	178	95	9	499	781
TOTAL	2892	1458	170	6573	11093

Tabla 10. Frecuencias Observadas. Factores Ubicación y Distrito.

FRECUENCIAS ESPERADAS					
DISTRITO	UBICACIÓN				TOTAL
	En acera	En calzada	En terrizo	No hay datos	
SE_BELLAVISTA-LA PALMERA	203.8712702	102.781574	11.98413414	463.3630217	782
SE_CASCO ANTIGUO	355.0801406	179.0134319	20.87262237	807.0338051	1362
SE_CERRO-AMATE	276.0865411	139.1888578	16.22915352	627.4954476	1059
SE_ESTE	269.8296223	136.0344361	15.86135401	613.2745876	1035
SE_LOCAL	129.3096547	65.19138195	7.60118994	293.8977734	496
SE_LOS REMEDIOS	136.6093933	68.87154061	8.030289372	310.4887767	524
SE_MACARENA	266.7011629	134.4572253	15.67745425	606.1641576	1023
SE_NERVION	276.607951	139.4517263	16.25980348	628.6805192	1061
SE_NORTE	265.1369332	133.6686199	15.58550437	602.6089426	1017
SE_SAN PABLO-SANTA JUSTA	243.7591274	122.8910124	14.32885604	554.0210042	935
SE_SUR	265.3976382	133.8000541	15.60082935	603.2014784	1018
SE_TRIANA	203.6105652	102.6501397	11.96880916	462.7704859	781
TOTAL	2892	1458	170	6573	11093

Tabla 11. Frecuencias Esperadas. Factores Ubicación y Distrito.

TÉRMINOS SUMATORIO COEFICIENTE DE CONTINGENCIA CHI CUADRADO				
DISTRITO	UBICACIÓN			
	En acera	En calzada	En terrizo	No hay datos
SE_BELLAVISTA-LA PALMERA	1.936846608	9.827329985	12.0476816	3.390624571
SE_CASCO ANTIGUO	31.09674322	125.6663837	7.938839875	1.271526252
SE_CERRO-AMATE	1.610517529	0.004727042	3.220171684	1.205587079
SE_ESTE	1.507781591	5.372615624	2.375772876	0.000858055
SE_LOCAL	11.94998903	15.89604394	0.890148668	18.68384346
SE_LOS REMEDIOS	4.719169688	0.899662633	31.75871326	3.612040916
SE_MACARENA	9.486171663	0.221492803	2.844109417	2.402819278
SE_NERVION	9.923528203	2.441462813	2.409939312	1.218760757
SE_NORTE	2.383166325	0.053277515	1.250377991	0.90795461
SE_SAN PABLO-SANTA JUSTA	16.93019564	0.194660304	0.773354305	5.66468219
SE_SUR	2.099006942	1.224524054	0.740624799	0.334352611
SE_TRIANA	3.22135077	0.570136953	0.736399729	2.836347028

Tabla 12. Términos Sumatorio Coeficiente de Contingencia Chi cuadrado. Factores Ubicación y Distrito.

$$\chi^2 = 367,7523$$

$$n (\text{grados de libertad}) = (12 - 1) \times (4 - 1) = 33$$

$$\text{Nivel de riesgo o de significación (5\%)} = 0.05$$

$$\chi^2_{7,0.05} = 47,3999 \quad \text{¿} \chi^2 > \chi^2_{7,0.05} \text{?} \rightarrow \text{Sí}$$

Debido a que el coeficiente χ^2 es mayor que el valor teórico, se concluye con la existencia de una relación entre estos dos factores.

Finalmente se calcula el coeficiente de K. Pearson y se cuantifica la relación.

$$C = \sqrt{\frac{\chi^2}{N + \chi^2}} = \sqrt{\frac{367.7523}{367.7523 + 11093}} = 0,1791312$$

Al igual que ocurría con la relación entre los factores Material y Ubicación, según el coeficiente χ^2 no existe independencia, aunque según muestra el coeficiente K. Pearson, la relación es muy baja.

2.2 Efecto de los factores sobre la aparición de incidencias.

2.2.1 Análisis de distribución de los factores.

Este subapartado está destinado a estudiar la distribución de los factores frente a las **dos redes principales que conforma la red**, y frente a los distintos municipios sobre los que se dispone la red de abastecimiento de EMASESA pertenecientes al área metropolitana de Sevilla. Para ello se hace uso de la macro o Software “Análisis de Factores”.

Todos los factores se van a presentar con relación al tipo de red y municipio.

Se realizará el estudio para cada uno de los factores, tanto cuantitativos como cualitativos, estableciendo un pequeño apartado para cada uno de ellos.

2.2.1.1 Material

En primer lugar, se realiza el estudio del factor Material. Este factor engloba todos los tipos de materiales que conforman la red de abastecimiento de EMASESA.

Cabe destacar con respecto a este factor, que tras el tratamiento de los datos realizado mediante el software “Genera Datos” con el objetivo de introducir los registros en los softwares que posteriormente serán utilizados, ciertos materiales con muy poca representación en términos de longitud e incidencias se agrupan en una sola clasificación llamada VA (Varios). El procedimiento y justificación se describen en [2].

La definición de las siglas correspondiente a los distintos materiales que aparecen a continuación se describe en 1.3.2 Factores considerados.

Para el caso el material se han obtenido los siguientes resultados de longitud instalada según el tipo de red, que se muestran en la tabla 13.

	AI	FC	FD	HACH	HF	PE	PVC	VA	TODOS
RSA	0	1039.027	1791.37	2.944	164.819	226.786	29.626	2.376	3256.947
RAA	0.397	77.639	253.256	129.219	16.271	4.646	0	2.302	483.73
TODOS	0.397	1116.667	2044.626	132.163	181.09	231.432	29.626	4.677	3740.678

Tabla 13. Longitud frente a Tipo de Red y Material.

En la figura 17 se representan gráficamente los resultados de longitud por material y tipo de red que aparecen en la tabla 13, de forma gráfica.

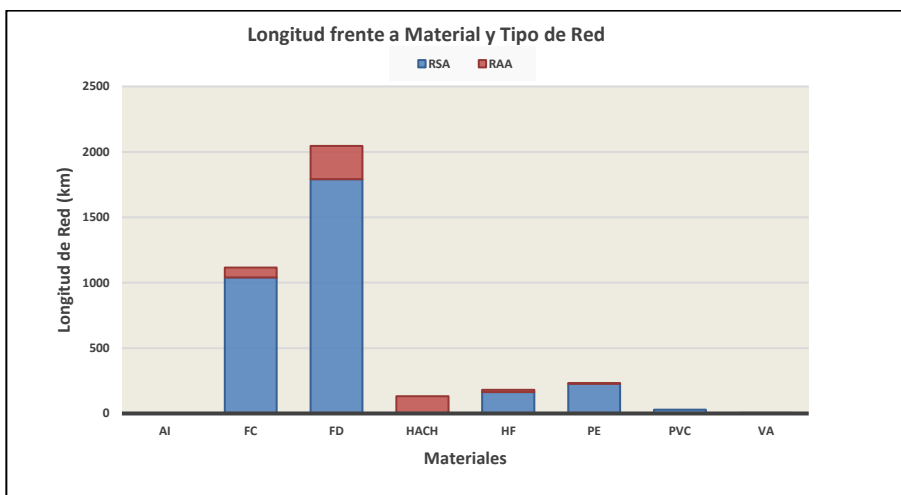


Figura 17. Longitud frente a Tipo de Red y Material.

En la tabla 14 se muestran los resultados de longitud de red instalada por tipo de material y municipio, que posteriormente aparecen representados de forma gráfica en la figura 18.

	AI	FC	FD	HACH	HF	PE	PVC	VA	TODOS
Alcalá de Guadaíra	0	189.041	269.55	21.7	10.296	18.08	9.605	0.703	518.975
Alcalá del Río	0	15.805	73.461	0	0.003	9.6	7.89	0.394	107.155
Burguillos	0	0	0.838	0	0	0	0	0.017	0.855
Carmona	0	0	0	0	0	8.959	0	0	8.959
Camas	0	41.567	50.439	5.487	0.257	21.188	0.317	0.12	119.376
Coria del Río	0	53.419	49.857	0	0	12.545	1.807	0.014	117.641
Dos Hermanas	0	197.169	450.24	6.646	5.704	18.136	3.562	0.393	681.85
El Garrobo	0	1.929	6.09	0	0	3.405	0	0.024	11.447
El Ronquillo	0	4.534	0.106	0	0	18.982	1.345	0.004	24.971
Gelves	0	3.31	3.402	0	0	0	0	0.047	6.759
Guillena	0	0	0	0	0	0	0	0	0
La Puebla del Río	0	21.386	23.774	0	0	2.9	1.62	0.043	49.724
Mairena del Alcor	0	39.979	100.209	0	0.014	3.894	1.393	0.365	145.854
Palomares	0	0.889	2.098	0	0	0.015	0	0.003	3.005
La Rinconada	0	99.029	119.583	0	0.006	36.459	0.739	0.056	255.873
Santiponce	0	0	1.804	0	0	0.247	0	0	2.051
Sevilla	0.397	430	844.738	96.193	164.644	84.217	1.347	2.364	1623.899
San Juan de Aznalfarache	0	18.609	48.437	2.137	0.165	1.765	0	0.132	71.244
TODOS	0.397	1116.667	2044.626	132.163	181.09	231.432	29.626	4.677	3740.678

Tabla 14. Longitud frente a Material y Municipio.

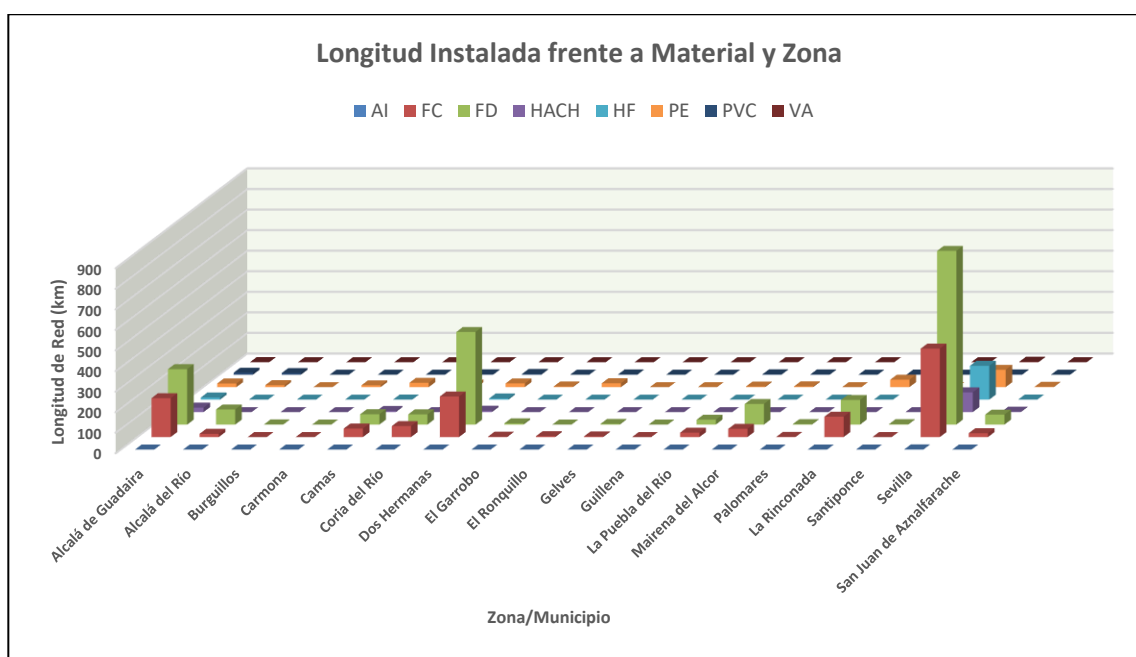


Figura 18. Longitud frente a Material y Municipio.

De la interpretación de la distribución de los distintos materiales sobre la red de abastecimiento y los diferentes municipios que conforman el área metropolitana de Sevilla, se extraen la siguiente información:

- **FD: Fundición Dúctil.** La longitud conformada por los tramos compuestos de este material es igual a 1791.37 km sobre la RSA, y de 253.256 km sobre la RAA, equivalente a un 55% de la RSA, a un 52.35% de la RAA. En cuanto a la longitud instalada de este material en los distintos municipios, se encuentra el predominio en el municipio de Sevilla, con una longitud igual a 844.738 km, equivalente a un 41.32% de la longitud total instalada de este material en la red de abastecimiento, y a un 52.02 % de la longitud total instalada en el municipio de Sevilla. Le siguen el municipio de Dos Hermanas, con una longitud total de FD instalada de 450.24 km, y Alcalá de Guadaíra, con una longitud de 265.55 km, respectivamente.
- **FC: Fibrocemento.** De este material está constituida una longitud igual a 1039.027 km de la RSA, equivalente a un 31.9% de la longitud total de dicha red, e igual a 77.639 km para la RAA, lo que supone un 16.05% de la longitud total de la RAA. En cuanto a la distribución de fibrocemento en los distintos municipios, el municipio donde más longitud hay instalada, es Sevilla, con una longitud igual a 430 km, equivalente al 38.51% de la longitud instalada existente de este material, y a un 26.48% de la longitud instalada en el municipio de Sevilla. A este municipio le sigue el municipio de Dos hermanas con una longitud de 197.169 km, y Alcalá de Guadaíra, con una longitud total de 189.041 km de longitud instalada de Fibrocemento.
- **PE: Polietileno.** Constituye 226.786 km de la longitud de la RSA, que supone un 6.96% de la longitud total de esta, y 4.646 km, equivalentes a un 0.96% de la longitud de la RAA. En referencia a la longitud instalada por municipio, destaca de nuevo Sevilla con 84.217 km de red instalados, seguido de La Rinconada, con 36.459 km, Camas, con 21.188 km, El Ronquillo, con 18.982 km, Dos Hermanas, con 18.136 km, y Alcalá de Guadaíra, con 18.08 km, respectivamente.
- **HF: Hierro Fundido.** Supone 164.819 km de la RSA, equivalente a un 5.06% de la longitud total de la RSA, y a 16.271 km de la RAA, equivalente a un 3.36% de la longitud total de la RAA. En cuanto a la distribución por municipios, destaca Sevilla con 164.644 km de red instalados de este material, correspondiente al 91% de longitud instalada de este material en la red de abastecimiento en el área metropolitana de Sevilla, seguido de Alcalá de Guadaíra, con una longitud de 10.296 km y Dos Hermanas, con una longitud igual a 5.704 km.
- **HACH: Hormigón Armado con Chapa.** Este material constituye un total de 2.944 km para la RSA y 129.219 km para la longitud total de la red, es decir, un 0.09% de la RSA y un 26.55% de la RAA. En cuanto a la distribución de longitud de red instalada en los distintos municipios, destacan el municipio de Sevilla con una longitud total de 96.193 km, seguido de Alcalá de Guadaíra con una longitud de 21.7 km, y Dos Hermanas y Camas, con una longitud de 6.646 km y 5.487 km respectivamente.
- **PVC: Policloruro de Vinilo.** Solamente existen tramos de PVC en la RSA con una longitud total de 29.262 km, que se corresponden con un 0.9% de la longitud de esta, y con un 0.78% de la longitud total de la red. En cuanto a los municipios que tienen más red instalada de PVC destacan Alcalá de Guadaíra, con una longitud de 9.605 km, seguido de Alcalá del Río con 7.89 km, y Dos Hermanas con 3.562 km.

- Al: **Aluminio**. No existen tramos de aluminio registrados en las bases de datos de EMASESA pertenecientes a la RSA. En la RAA existe una longitud igual a 0.397 km, que se corresponde con el 0.01% de la longitud total. Toda la longitud de instalada con dicho material se encuentra localizada en el municipio de Sevilla.
- VA: **Varios**. Este grupo abarca en torno a 2.376 km de la red de abastecimiento secundaria y 2.302 de la RAA, es decir, un 0.07% y 0.476% respectivamente. El municipio que más longitud instalada en su red en disposición de este grupo de materiales es Santiponce 2.364 km.

2.2.1.2 Diámetro

En cuanto al factor diámetro, sólo se ha representado la longitud de red instalada frente a diámetro y tipo de red, en base a que, tras observar los registros se ha comprobado que la variabilidad de diámetros está más asociada al tipo de red.

La distribución en función de la longitud de dicho factor sobre los dos tipos de redes se muestra en la tabla 15, y de forma gráfica en la figura 19.

	Diámetro	20	25	32	40	50	60	63
Tipo de Red	RSA	2481.275	2262.033	4921.815	3220.892	33877.412	83067.718	34612.47
	RAA	0.3	17.647	6115.548	1.01	107.449	10677.813	3132.045
	TOTAL	2481.575	2279.68	11037.363	3221.902	33984.861	93745.531	37744.515

70	75	80	90	100	110	125	150	160
10801.567	23763.602	412986.424	76648.832	1238648.923	74640.658	85444.757	733799.199	39688.185
16287.704	25823.047	70646.343	452.337	14470.8	85.052	64359.574	6501.036	62318.677
27089.271	49586.649	483632.767	77101.169	1253119.723	74725.71	149804.331	740300.235	102006.862

175	200	250	300	350	400	450	500	533
4120.21	412421.626	114383.372	66079.979	4262.253	4217.894	8.777	2420.42	124.25
8493.766	194.863	78946.409	739.766	36758.749	1824.575	45261.557	4655.871	556.679
12613.976	412616.489	193329.781	66819.745	41021.002	6042.469	45270.334	7076.291	680.929

600	800	1000	1300	1700	Total
10306.625	635.085	393.159	299.748	2.655	3480541.815
24487.034	266.775	2.212	23248.006	368.792	506801.436
34793.659	901.86	395.371	23547.754	371.447	3987343.251

Tabla 15. Longitud por Tipo de Red y Diámetro.

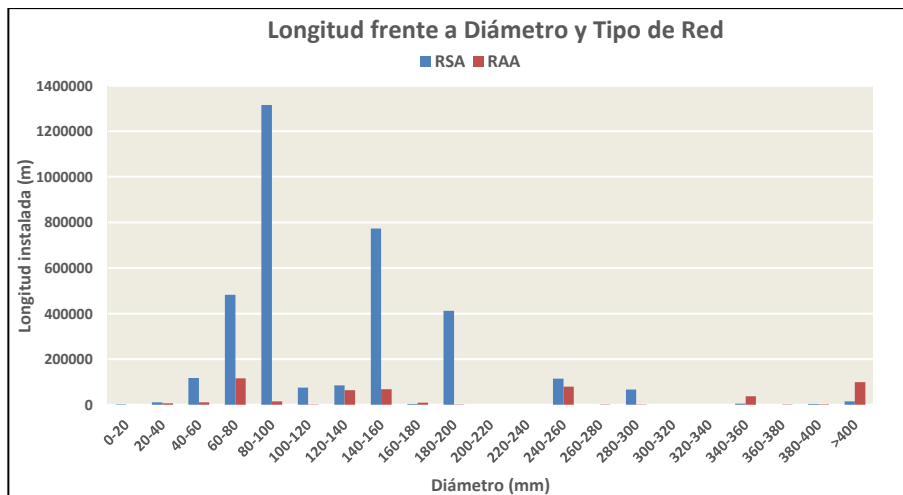


Figura 19. Longitud por Tipo de Red y Diámetro.

Como se puede observar tanto en la figura 18, como en la tabla 6 la mayor longitud la conforman las familias formadas por tramos de diámetro 100 mm, en la RSA con una longitud total de 1238.64 km de longitud, equivalente al 35.6% de la longitud total instalada en la RSA, y a un 98.84% de la longitud instalada de tuberías con dicho diámetro.

Por otro lado, se encuentran las tuberías con diámetro 150 mm, pertenecientes a la RSA con una longitud total de 733,8 km, correspondientes a un 21.08 % de la longitud total instalada sobre la RSA, y a un

99.121% de la longitud de tuberías instaladas con dicho diámetro.

Finalmente, los dos diámetros que más destacan respecto a los demás son diámetro 200, con una longitud de 412.421 km correspondientes a la RSA, equivalentes a un 11.84% de la longitud instalada en la RSA, y a un 99.95% de la longitud de tuberías con dicho diámetro, y el diámetro 80, con una longitud de 412.986 km correspondientes a la RSA, equivalentes a un 11.86 % de la longitud instalada sobre la RSA y a un 85.9 % de la longitud conformada por las familias con dicho diámetro.

2.2.1.3 Ubicación

La distribución de la red de abastecimiento en función de la ubicación, y el tipo de red frente a las distintas zonas/municipios se muestra en la tabla 16, y posteriormente de forma gráfica en la figura 20.

	En acera	En calzada	En terrizo	No hay datos	TODOS
RSA	845.098	225.048	36.19	2150.612	3256.947
RAA	50.775	40.503	87.388	305.065	483.73
TODOS	895.872	265.551	123.578	2455.677	3740.678

Tabla 16. Longitud frente a Tipo de Red y Ubicación.

En referencia a la distribución de la longitud frente a las dos tipologías de redes encontradas, se observa que el mayor porcentaje de red está incluida dentro de la RSA, que cuenta con un total de 3256.947 km, que supone el 87% de la longitud total de la red, frente a los 483.73 km restantes, pertenecientes a la Red Arterial y de Transporte (RAA).

Dentro de la RSA, la mayor parte de la longitud se encuentra bajo la ubicación “No hay Datos”, concretamente, una longitud igual a 2150.612 km de longitud, equivalente a un 66% de la longitud total de la RSA.

Por otro lado, está registrada bajo la referencia “En Acera” una longitud igual a 845.098 km, equivalente al 26 % de la longitud total de la RSA.

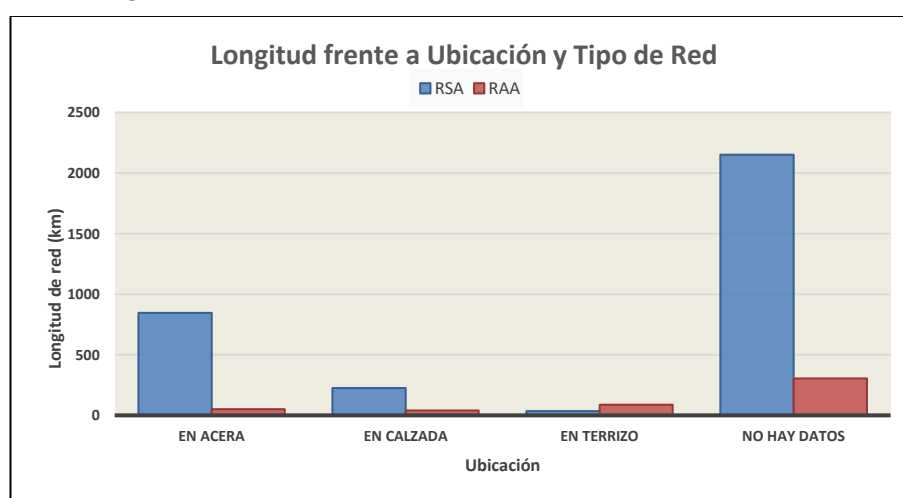


Figura 20. Longitud frente a Tipo de Red y Ubicación.

En la tabla 17 se representa los resultados en términos de longitud instalada por tipo de ubicación y municipio.

	En acera	En calzada	En terrizo	No hay datos	TODOS
Alcalá de Guadaira	166.171	42.841	22.576	287.386	518.975
Alcalá del Río	20.027	12.597	17.928	56.602	107.155
Burguillos	0	0	0.855	0	0.855
Carmona	1.248	4.772	0	2.939	8.959
Camas	14.183	12.104	3.569	89.52	119.376
Coria del Río	23.35	7.746	0.266	86.279	117.641
Dos Hermanas	191.623	33.294	26.807	430.126	681.85
El Garrobo	2.355	3.775	0.831	4.486	11.447
El Ronquillo	7.788	2.74	3.221	11.222	24.971
Gelves	0	3.136	0.283	3.34	6.759
Guillena	0	0	0	0	0

La Puebla del Río	9.326	3.814	0.176	36.407	49.724
Mairena del Alcor	40.148	15.486	1.762	88.458	145.854
Palomares	0.007	0	0	2.998	3.005
La Rinconada	76.429	22.344	11.443	145.655	255.873
Santiponce	0.068	0	0	1.983	2.051
Sevilla	337.75	100.522	31.676	1153.95	1623.899
San Juan de Aznalfarache	6.648	5.151	2.184	57.262	71.244
TODOS	895.872	265.551	123.578	2455.677	3740.678

Tabla 17. Longitud frente a Ubicación y Municipio.

Los resultados de la tabla 17 se muestran de forma gráfica en la figura 21.

En cuanto a la distribución de la longitud de red en función del factor Ubicación y el factor Municipio/Zona, los municipios que más longitud instalada tienen son: Sevilla, con 1153.95 km de RSA bajo la ubicación No Hay Datos, 337.75 km bajo la ubicación En Acera, 100.522 km bajo la ubicación En Calzada, y 31.676 km bajo la ubicación En Terrizo; Dos Hermanas, con 430.126 km, bajo la ubicación No Hay Datos y 191.623 km, bajo la ubicación En Acera; y Alcalá de Guadaíra, con 287.386 km registrados con el valor de ubicación No hay Datos, y 166.171 km registrados bajo la ubicación En Acera.

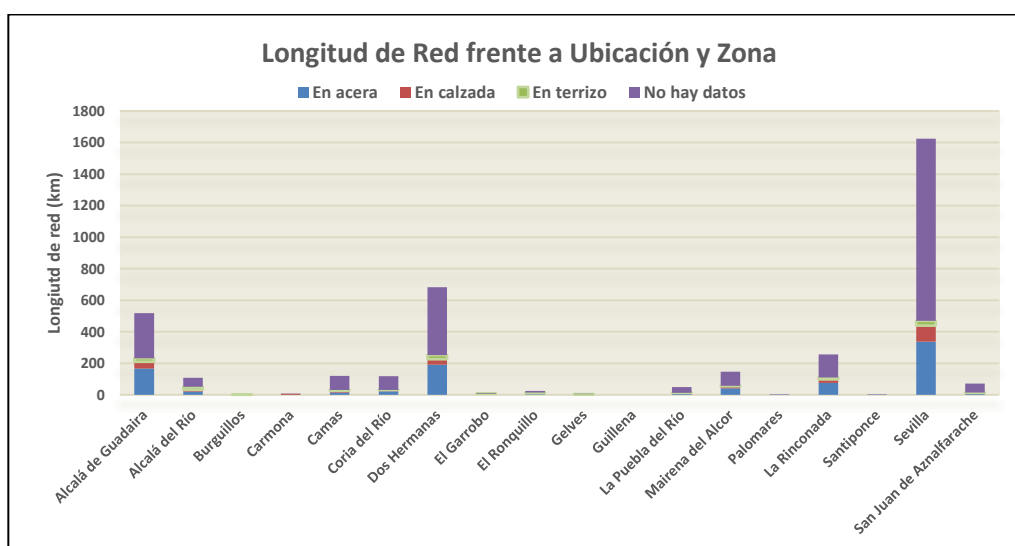


Figura 21. Longitud frente a Ubicación y Municipio.

2.2.1.4 Distrito

El distrito es un factor que representa una localización determinada de un tramo o familia, dentro de un mismo municipio.

Se presentan en la figura 22, la distribución de longitud de la Red de abastecimiento en función de los distritos en los que se localice.

Los cuatro distritos que más longitud disponen son, AG_ALCALA DE GUADAIRA, con 273.354 km pertenecientes a la RSA, seguido de SE_ESTES, con 201.282 km, SE_NORTE, con 194.874 km, y SE_CERRO-AMATE con 186.787 km.

De los cuatro distritos con mayor longitud de red instalada, tres de ellos pertenecen al municipio de Sevilla.

La tabla con las longitudes de red en función del distrito se muestra en la tabla 44 (Anexo).

En cuanto a la distribución por tipo de red, simplificando el resultado a los distritos pertenecientes municipio de Sevilla, se muestra en la tabla 18.

DISTRITOS	RSA	RAA	TODOS
SE_BELLAVISTA-LA PALMERA	131.077	29.572	160.649
SE_CASCO ANTIGUO	125.923	13.736	139.659
SE_CERRO-AMATE	186.787	10.937	197.723
SE_ESTE	201.282	28.018	229.3
SE_LOCAL	0.367	0	0.367
SE_LOS REMEDIOS	53.181	10.698	63.879
SE_MACARENA	102.965	8.901	111.866
SE_NERVION	97.077	5.92	102.997
SE_NORTE	194.874	23.16	218.034
SE_SAN PABLO-SANTA JUSTA	125.597	9.86	135.457
SE_SUR	140.856	21.602	162.457
SE_TRIANA	81.451	20.058	101.509

Tabla 18. Longitud frente a Tipo de Red y Distrito.

Se puede apreciar en la figura 23, que la mayoría de distritos pertenecientes al municipio de Sevilla, consta de una longitud de red considerable.

La mayor parte de la Red de Abastecimiento en términos de longitud se distribuye en la RSA, destacando los distritos de: SE_ESTE, con 201.282 km, equivalente al 13.96 % de la longitud total de la RSA, SE_NORTE, con 194.874 km, equivalentes al 13.52% de la longitud de la RSA, y SE_CERRO-AMATE, con 186.787 km, equivalentes al 12.96% de la longitud total de la RSA.

No obstante, la mayoría de distritos cuentan con una gran longitud de red, con excepción el distrito SE_LOCAL; que cuenta con una longitud muy inferior a los demás. Además, este distrito se trata de un distrito auxiliar que se generó para agrupar a todos aquellos tramos que no tenían un distrito establecido, lo cual muestra que la mayoría de los tramos tenían distrito asignado, dada la poca longitud que representa (véase documento [2]).

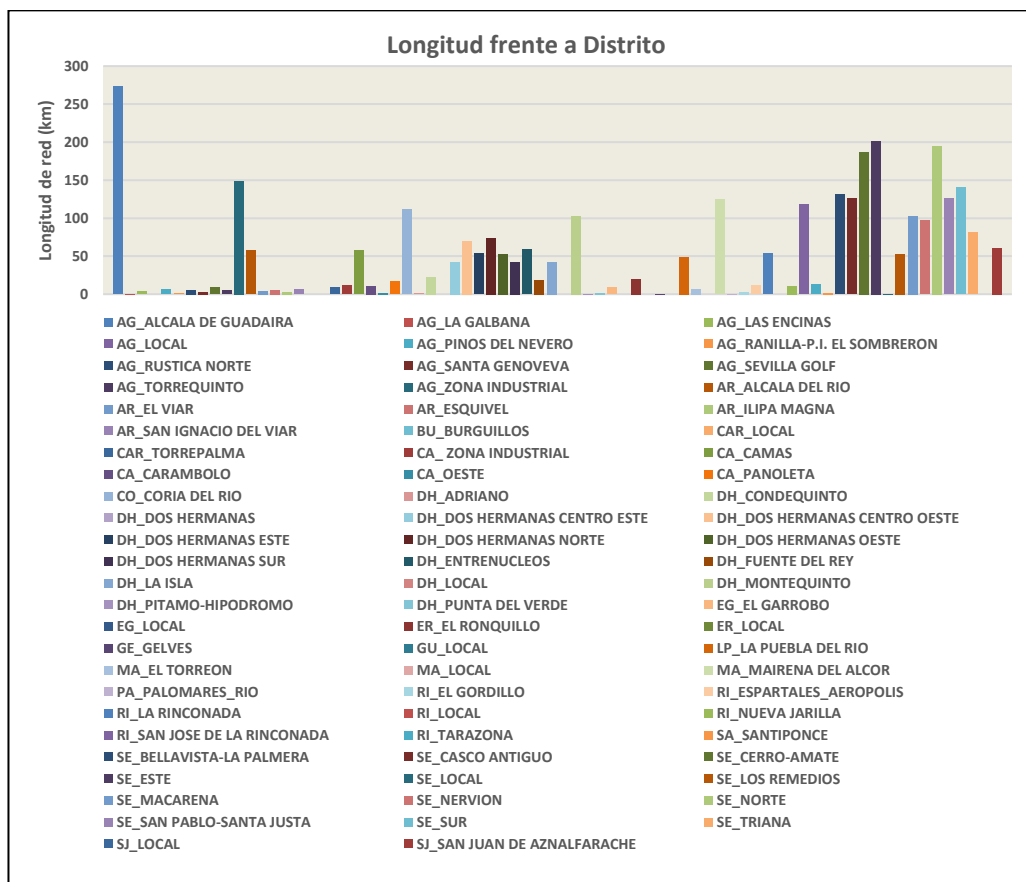


Figura 22. Longitud frente a Distrito.

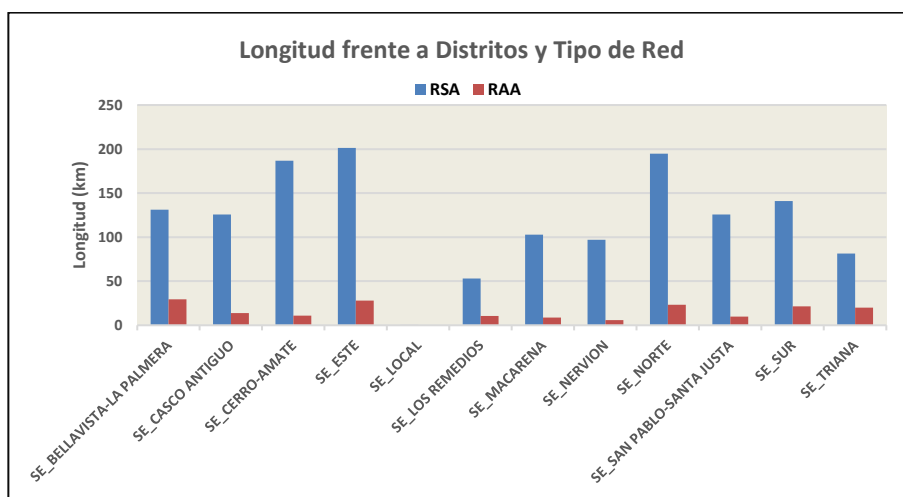


Figura 23. Longitud frente a Distrito y Tipo de Red.

2.2.1.5 Año de Instalación

Aunque se va a realizar el estudio para el factor Año de Instalación, este no es un factor cuantitativo como otro, por lo que requiere un especial tratamiento, debido a que en el procedimiento de regresión utilizado (véase [2]) se introduce para el cálculo de la edad, que es el factor explícitamente considerado en el desarrollo.

Aún así, como se trata de un factor que se define explícitamente en las BBDD de EMASESA, y ha sido utilizado (aunque no directamente) en el estudio, también va a disponer de un análisis de red y de históricos de incidencias.

En la tabla 19 se muestra la distribución de la red, en términos de longitud, frente al año de instalación de los distintos tramos que conforman las familias, y los dos tipos de red existentes, que aparecen posteriormente representados gráficamente en la figura 24.

LONGITUD VS TIPO DE RED		Año Instalación								
		1900	1910	1950	1955	1960	1962	1963	1965	
RSA		124.25	0	69828.177	1.351	153524.924	236.511	25.528	47246.68	
RAA		0	8493.766	8568.774	0	31443.861	0	0	4534.691	
1966	1967	1968	1969	1970	1971	1972	1973	1974	1975	
3.775	162.599	3071.778	628.147	296417.61	3850.075	2467.794	330.801	1641.382	148550.965	
0	0	1.495	0	21598.559	1.709	1601.589	0	36.79	58183.327	
1976	1977	1978	1979	1980	1981	1982	1983	1984	1985	
1280.281	462.966	1667.426	3105.445	150182.639	1936.874	820.853	716.351	571.429	114487.799	
79.499	195.972	2	321.982	13777.867	435.548	0	0	0	18134.441	
1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	
1586.65	2925.961	10594.159	2347.316	224288.65	44512.36	45188.853	34000.811	33459.293	127209.283	
483.539	240.427	54.356	506.543	28268.844	29938.802	25678.696	1924.611	68.633	8548.613	
1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	
46219.491	62950.15	60595.862	83656.748	148802.24	99178.935	113528.241	137812.993	104633.398	107658.215	
5764.727	12053.693	10358.858	10459.191	9790.392	11550.664	32185.216	12018.972	2547.491	5124.424	
2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	
128167.211	157388.272	119554.811	122166.701	128869.57	60485.82	51904.854	23008.187	21323.793	24117.899	
38158.458	12395.858	9974.017	9709.044	8267.915	7770.685	6853.91	655.897	5436.664	1314.734	
2016	2017	2018								
24718.142	31593.006	51681.946								
2139.427	5865.044	4401.758								

Tabla 19. Longitud frente a Tipo de Red y Año de Instalación.

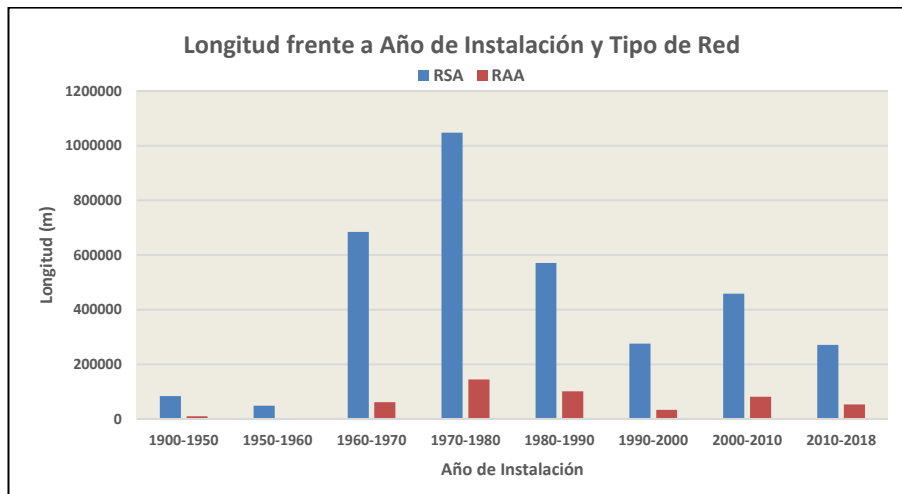


Figura 24. Longitud frente a Tipo de Red y Año de Instalación.

En referencia a la RAA, los años de instalación que más longitud instalada tienen registrada son: 1975, con 58183.327 m, 2006, con 38158.458 m, 2002, con 32185.216 m, 1960, con 31443.861 m, y 1991, con 29938.802 m.

Dado que el histórico de incidencias comienza a desarrollarse a partir de 2003, se muestra en la figura 25 la longitud de red instalada anualmente para el periodo 2003-2018.

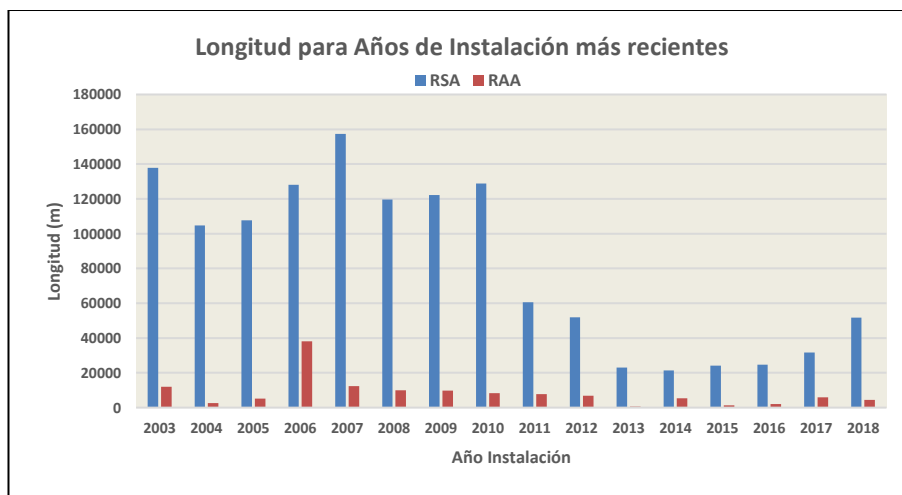


Figura 25. Longitud frente a Año de Instalación. Años de Instalación más recientes.

De entre los Años de Instalación más recientes, destaca los que pertenecen a la primera década del siglo XX, concretamente, el año 2007, con una longitud instalada en la RSA, de 157388.272 m, y una longitud en la RAA de 12395.858 m; el año 2003, con una longitud instalada en la RSA de 137812.993 m, y 12018.972 m en la RAA, y el año 2010, con 128869.57 m en la RSA y 38158.458 m en la RAA.

2.2.1.6 Número de Acometidas

Las acometidas de una tubería son el número de enlaces adicionales que posee.

Existe un amplio rango de valores en cuanto al número de acometidas de que disponen los tramos sobre los que se posee información. Debido a este motivo, para representar la distribución de la red de abastecimiento en términos de longitud frente a número de acometidas, se ha optado por utilizar un gráfico de dispersión. Los distintos valores de acometidas que se pueden encontrar van desde 0 (12374 familias), hasta 912 acometidas (1 familia). En la figura 25, se representan las familias existentes, frente al número de acometidas que poseen y de su longitud. Los valores de longitud en función del número de acometidas se muestran en la tabla 45 (Anexo).

Con idea de representar el número de acometidas por unidad de longitud, para RSA y RAA, se muestra la figura 27, en la que se representa en el eje de ordenadas el número de acometidas por unidad de longitud (km) y en el eje de abscisas la longitud, por familias. Para mejor visualización de los valores se han empleado escalas logarítmicas.

La mayor parte de las familias poseen un número de acometidas de entre 60 y 300, aunque las familias que representan más longitud no tienen acometidas.

En la figura 26 se puede apreciar que existe un gran número de familias con elevado índice de acometidas por km. Cuanto más aumenta la longitud de la familia, más se reduce el índice de acometidas por unidad de longitud

En el caso de las familias pertenecientes a la RAA, no poseen un gran número de acometidas, y en términos de índices de acometidas por unidad de longitud, también presentan valores inferiores a las familias pertenecientes a la RSA.

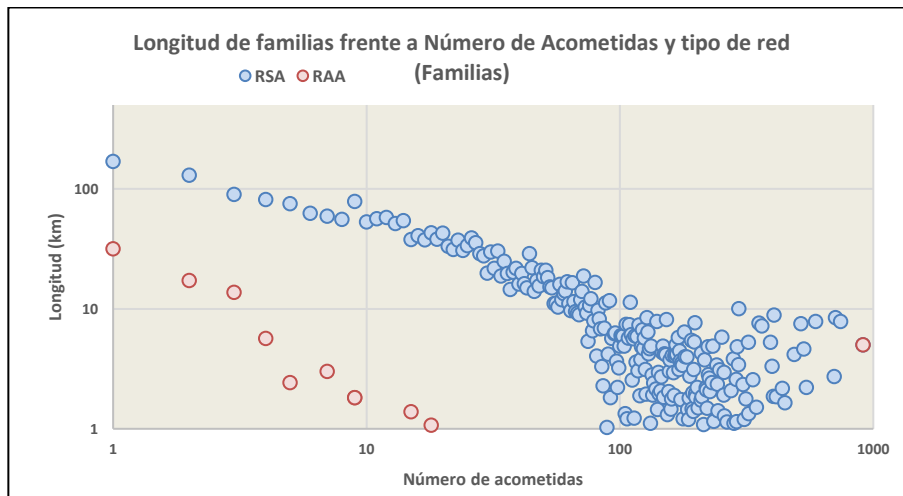


Figura 26. Longitud frente a factor Número de Acometidas y Tipo de Red. Familias.

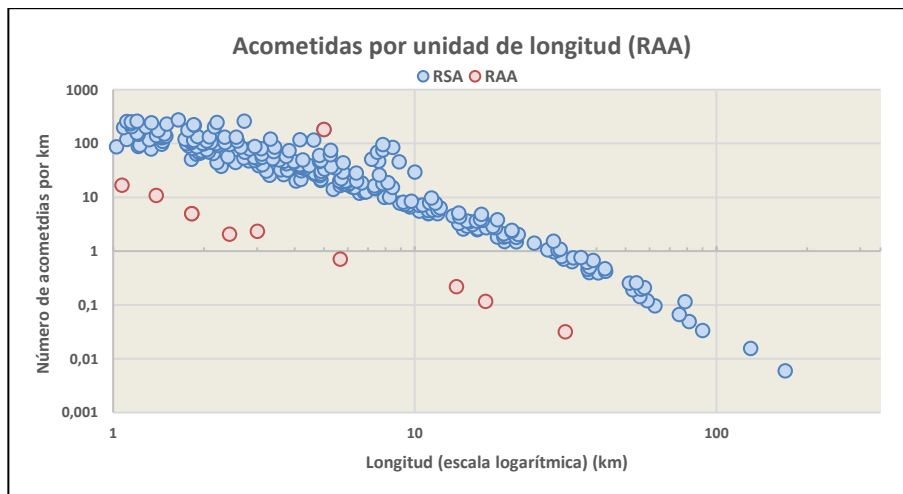


Figura 27. Número de Acometidas por unidad de longitud correspondiente a familias.

2.2.1.7 Presión Media

El factor Presión Media a la hora de generar las familias, se obtenía realizando una media de todas las presiones de los distintos tramos que conformaban la familia; de ahí proviene la clasificación del factor como “Factores que se promedian”.

Como se puede observar en la figura 28, la mayoría de las familias están conformadas por tramos sometidos a valores de presión que varían entre 25 y 80 m.c.a. Resultados en tablas 48 y 49 (Anexo).

Existen familias concretas que presentan una presión mucho mayor, pero con una longitud muy

pequeña.

La familia que más longitud presenta está conformada por tramos con una media de presión de 30 m.c.a, y está dotada de una longitud de 182413.152 metros.

En cuanto a la RSA, la mayor longitud de red está distribuida en familias que presentan rangos de valores de presión entre 25 y 35 m.c.a.

Las familias pertenecientes a la RSA, conformadas por tramos con un valor de presión media superior a 80 m.c.a, suponen el 0.18% de la longitud de la RSA.

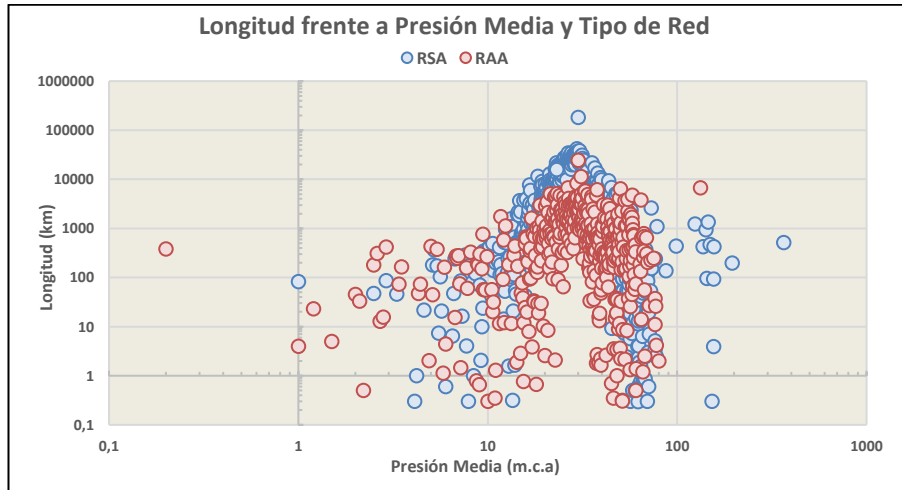


Figura 28. Longitud frente a factores Presión media y Tipo de Red.

2.2.1.8 Fluctuación de Presión

En la generación de las familias, este factor se clasifica dentro del grupo “Factores que se Promedian”, al igual que, la Presión Media.

A continuación, se representa en la figura 29 la longitud instalada en la Red de abastecimiento para el factor Fluctuación de Presión, y para los dos tipos de red existentes. Resultados en tablas 51 y 52 (Anexo).

La mayor parte de las familias que conforman la Red de Abastecimiento de EMASESA, sufren una fluctuación de presión de entre 0 y 7 m.c.a.

Las variaciones de presión en los tramos, por lo tanto, pueden llegar a suponer un 25% de la presión media general, que toma valores entre 25-35 m.c.a. Un 25% de fluctuación de presión respecto a la presión media se trata de una fluctuación considerable.

La familia que presenta la mayor longitud dispone de 505737 m, y presenta una fluctuación de presión de 3.1 m.c.a. respectivamente.

Existen algunas familias con valores de fluctuación de presión por encima de 10 m.c.a, pero estas presentan sólo el 0.93% de la longitud de la RSA.

Las familias con valores de fluctuación de presión elevados tienen más relevancia en el caso de la RAA, donde las familias que poseen valores por encima de 10 m.c.a, suponen un 61% de la longitud total de la RAA.

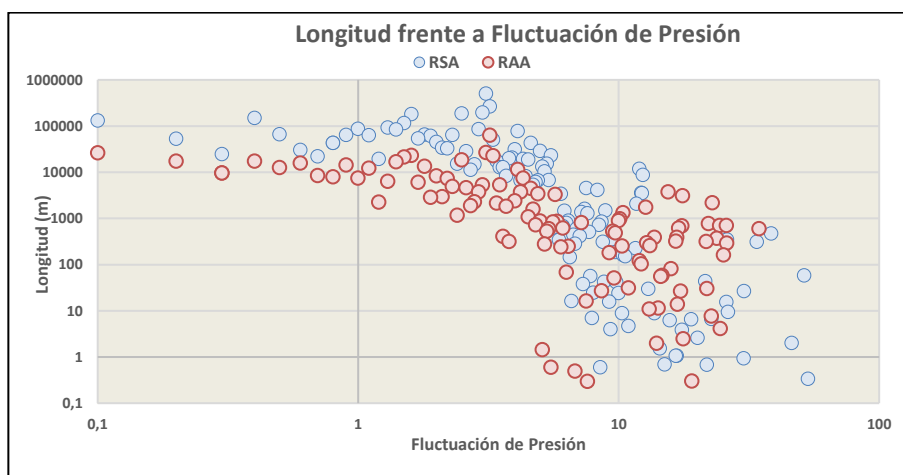


Figura 29. Longitud frente a Fluctuación de Presión y Tipo de Red.

2.2.2 Análisis del efecto de los factores sobre la aparición de incidencias.

Este análisis se va a realizar seleccionando los datos correspondientes al tipo de red y municipio que represente/suponga/etc. la mayor longitud. Para ello, en la tabla 20 se muestran las longitudes de red instaladas por municipio y tipo de red, que a su vez se representan de forma gráfica en la figura 30.

	RSA	RAA	TODOS
Alcalá de Guadaíra	454.89	64.085	518.975
Alcalá del Río	76.06	31.095	107.155
Burguillos	0	0.855	0.855
Carmona	8.959	0	8.959
Camas	107.137	12.239	119.376
Coria del Río	111.791	5.85	117.641
Dos Hermanas	583.715	98.135	681.85
El Garrobo	8.449	2.997	11.447
El Ronquillo	19.719	5.252	24.971
Gelves	0.028	6.731	6.759
Guillena	0	0	0
La Puebla del Río	48.704	1.019	49.724
Mairena del Alcor	131.185	14.669	145.854
Palomares	0.414	2.592	3.005
La Rinconada	211.386	44.487	255.873
Santiponce	1.518	0.532	2.051
Sevilla	1441.437	182.462	1623.899
San Juan de Aznalfarache	60.514	10.731	71.244
TODOS	3256.947	483.73	3740.678

Tabla 20. Longitud frente a Tipo de Red y Municipio.

Como puede observarse tanto en la tabla 20 como en la figura 30, la mayor parte de la longitud corresponde a la RSA perteneciente al municipio de Sevilla, correspondiente a 1441.437 km de red, equivalente a un 44.25% de la longitud total de la RSA, y a un 38.53% de la longitud total de la red de abastecimiento. Le siguen la parte correspondiente a la RSA en el municipio de Dos Hermanas (583.715 km) y Alcalá de Guadaíra (454.89 km).

En cuanto a la RAA, la mayor parte de la longitud se encuentra también en el municipio de Sevilla (182.462 km).

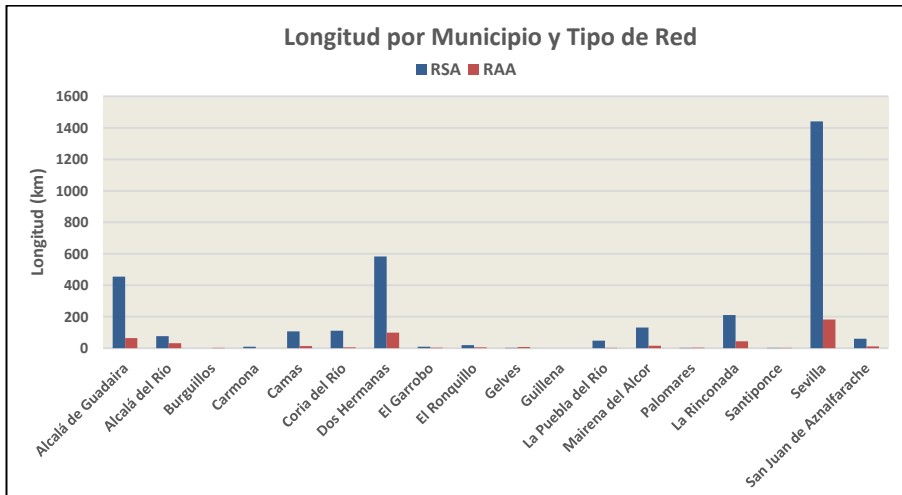


Figura 30. Longitud red frente a Tipo de Red y Municipio.

La RSA está constituida por 1441.437 km, de los 1623.899 km totales que existe en la red de abastecimiento dentro del municipio de Sevilla. Esta cantidad equivale a un 88.7 % de la longitud de la red, dato que justifica aún más la decisión de elegir la RSA como objeto de estudio.

A la vista de los resultados, el análisis se realizará con los datos de la RSA y del municipio de Sevilla.

Para realizar el estudio de incidencias es necesario, además, tomar un período de referencia para el análisis, es decir, definir el período sobre el que se van a estudiar las incidencias.

2.2.2.1 Material

Las consultas de incidencias registradas en función del factor material, acotando los resultados para la RSA en el Municipio de Sevilla, y registradas para un periodo que comprende desde el año 2005 hasta el año 2018, se muestran en la tabla 21, y posteriormente, se hallan representadas de forma gráfica en las figuras 31 y 32, respectivamente.

		Incidencias	Inc./Long	Long. (m)
Material	FC	2682	5.74664633	466706.988
	FD	186	0.23467256	792593.725
	HACH	20	2.62691031	7613.507
	HF	1745	9.18884175	189904.239
	PE	22	0.25017808	87937.359
	PVC	1	0.504464	1982.302
	VA	15	14.429945	1039.505

Tabla 21. Incidencias absolutas e incidencias por unidad de longitud en función del factor Material.

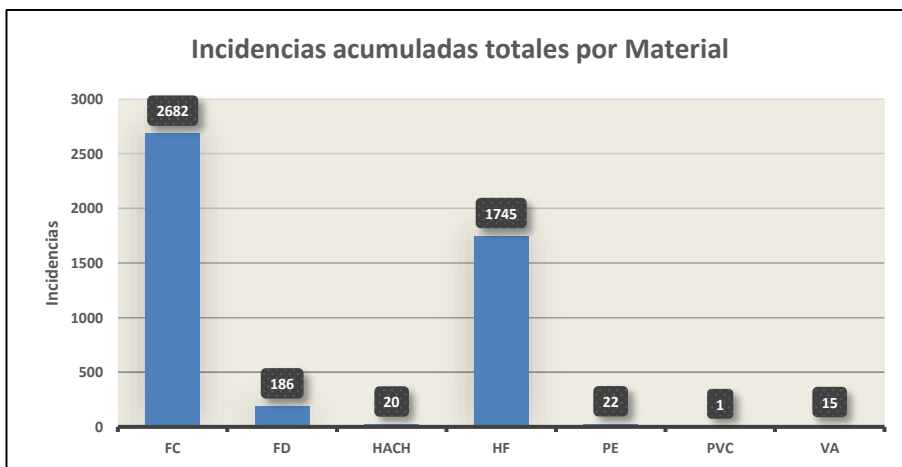


Figura 31. Incidencias acumuladas por Material.

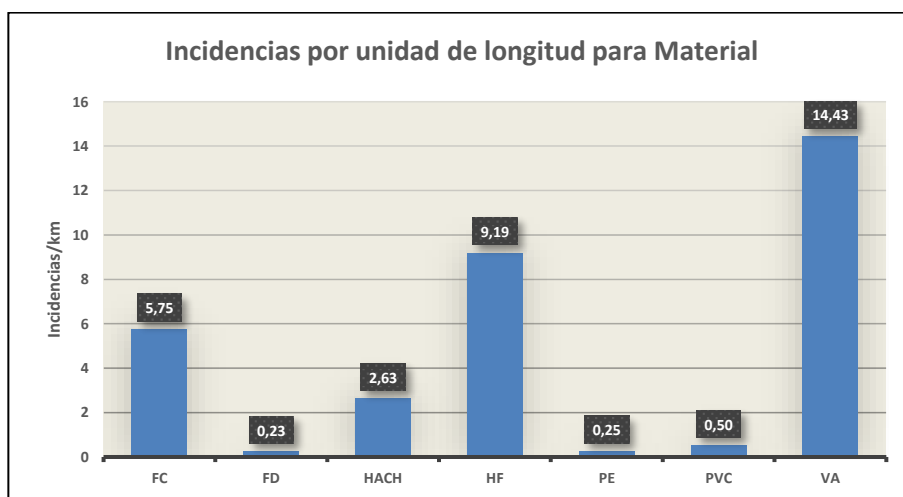


Figura 32. Incidencias por km para factor Material.

En términos de incidencias acumuladas, los materiales que más incidencias presentan en el periodo establecido son: FC con 2682 incidencias, equivalente al 57% de las incidencias totales en ese periodo, Hierro Fundido (HF) con 1745 incidencias, equivalentes al 37%, y Fundición Dúctil (FD) con 186 incidencias, equivalentes al 4%.

Relativo a la longitud, destacan: el grupo Varios (VA) con 14.43 incidencias por km equivalente a un 44% de las incidencias totales por unidad de longitud, Hierro Fundido (HF) con 9.19 incidencias por km, equivalente a un 28 %, y FC, con 5.75 incidencias por km, equivalentes a un 17%.

Es importante comentar que el grupo VA, no presenta un número muy elevado de incidencias acumuladas, como por ejemplo FC, pero sí tiene un alto índice de incidencias por unidad de longitud. Esto es debido a que este grupo de materiales dispone de una longitud muy pequeña. El hecho de presentar estas características clasifica a los registros definidos por este material como potenciales *outliers*, como ya se verá que pasa con alguno más, como PE o PVC.

2.2.2.2 Diámetro

En cuanto a los resultados de incidencias en relación con el factor diámetro, se representan en la tabla 22. A su vez, se muestran de forma gráfica las incidencias acumuladas y por unidad de longitud frente al factor diámetro, de forma gráfica en las figuras 33 y 34.

	Diámetro						
	20	25	32	40	50	60	63
Incidencias	0	0	1	0	3	18	2
Inc./Long	0	0	0.666143966	0	0.957914658	11.44888138	0.524525776
Long. (m)	745.44	624.979	1501.177	810.549	3131.803	1572.206	3812.968

	70	75	80	90	100	110	125	150
12	246	523	16	1993	2	527	800	
22.79215266	12.99111154	8.233810494	0.63425419	3.295448316	0.069384523	8.362116787	2.131654153	
526.497	18936.024	63518.586	25226.479	604773.557	28824.872	63022.32	375295.401	

	160	175	200	250	300	350	400	500
0	10	357	83	44	9	2	19	
0	3.209626826	1.683835643	1.681660119	0.804115316	2.627500943	1.078340929	8.001280205	
22944.326	3115.627	212015.942	49355.99	54718.52	3425.308	1854.701	2374.62	

	600	800	1000	1300
0	4	0	0	
0	7.347727532	0	0	
4691.055	544.386	160.238	254.054	

Tabla 22. Incidencias absolutas e incidencias por unidad de longitud en función del factor Diámetro.

Para una mayor comprensión de la distribución de incidencias en el periodo que abarca desde el año 2005 hasta el 2018, se muestran las incidencias acumuladas y por unidad de longitud frente al factor diámetro,

de forma gráfica en las figuras 33 y 34.

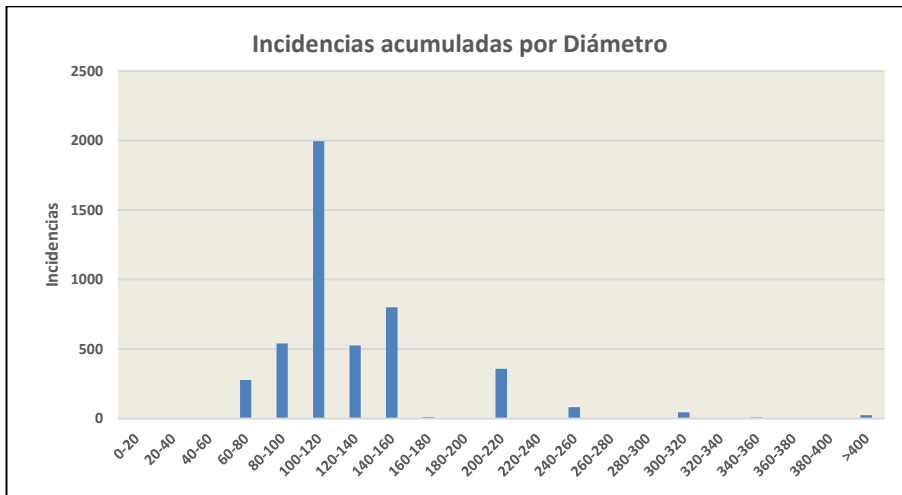


Figura 33. Incidencias acumuladas por Diámetro.

Las familias conformadas por los tramos con diámetro 100 mm, son las familias que más incidencias tienen contabilizadas desde el año 2005 al 2018, con un total de 1993 incidencias, equivalente a un 42.67 % de las incidencias totales. Con bastante diferencia, el segundo conjunto de tramos que más incidencias han registrado en este periodo son los que disponen de un diámetro de 150 mm, con un total de 800 incidencias, correspondiente a un 17.13% de las incidencias registradas.

De entre los valores que más destacan, se encuentran también los diámetros de valor 80 mm, 125 mm y 200 mm, con un total de incidencias registradas de 523, igual al 11.2%, 527 incidencias para diámetro 125, igual al 11.28% y, 357 incidencias para el diámetro 200, equivalente al 7.64% de las incidencias totales.

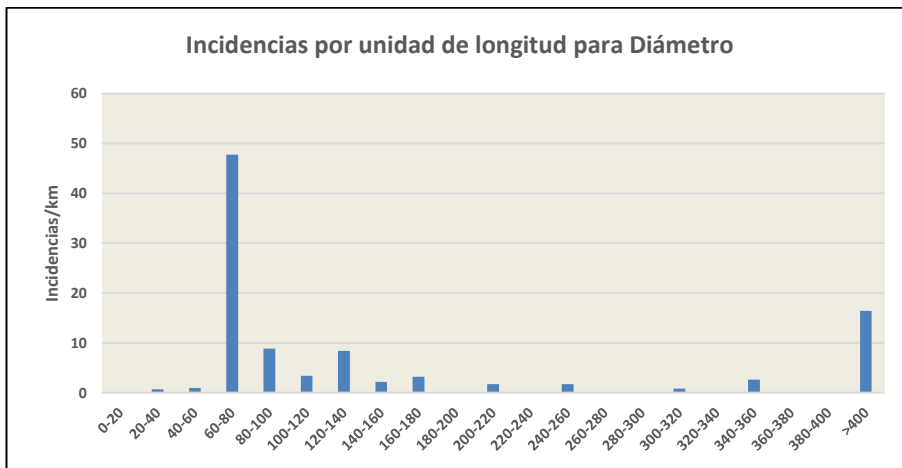


Figura 34. Incidencias por km para factor Diámetro.

En cuanto a los resultados de incidencias por unidad de longitud, las tuberías o tramos más afectados, son las que conforman las familias definidas por diámetro 70 mm, con un valor de 22.8 incidencias por unidad de longitud, seguidas de las familias con diámetro 75, con un índice de incidencias por unidad de longitud de 13, las de diámetro 60, con 11.45 incidencias por unidad de longitud, y las de diámetro 125, con un valor de 8.36 incidencias por unidad de longitud.

Destacan también las familias definidas por diámetros 80, 500 y 800 mm, con valores de 8.23, 8 y 7.35 incidencias por unidad de longitud, respectivamente.

2.2.2.3 Ubicación

Se muestran en la tabla 23 las incidencias acumuladas totales y por unidad de longitud para el periodo 2005-2018, para la Red Secundaria de Abastecimiento, en el municipio de Sevilla.

		Incidencias	Inc./Long	Long. (m)
Ubicación	En acera	1535	4.491477585	341758.357
	En calzada	336	3.641733459	92263.754
	En terrizo	71	3.300955762	21508.922
	No hay datos	2729	2.498520041	1092246.592

Tabla 23. Incidencias acumuladas y por unidad de longitud frente al factor Ubicación.

Los resultados de la tabla 23, se muestran en las figuras 35 y 36, respectivamente.

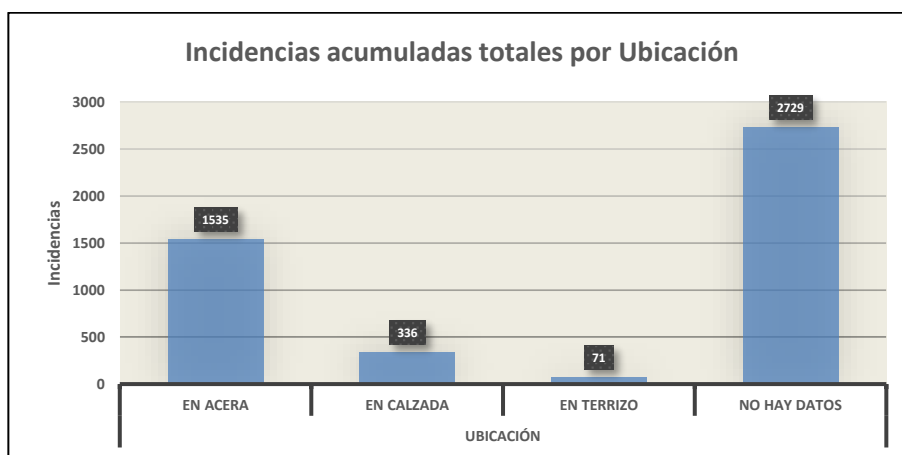


Figura 35. Incidencias acumuladas por Ubicación.

En términos de incidencias acumuladas, el 58% de las incidencias se dan en familias cuyos tramos no tienen asignada una ubicación. Este 58% es equivalente a 2729 incidencias.

Por otro lado, aproximadamente un tercio de las incidencias (33%) se dan en las familias cuyos tramos tienen asignada la ubicación “En acera”. Este 33 % equivale a 1535 incidencias registradas en el periodo 2005-2018.

Sólo un 7% de las incidencias se dan en los tramos que tienen asignada la ubicación de “En calzada”, y un 2% se dan en tramos con ubicación “En terrizo”.

En referencia a las incidencias por unidad de longitud, en los tramos con ubicación localizada “en Acera”, es donde se da la mayor tendencia a sufrir una incidencia, del orden de 4.49 incidencias por unidad de longitud.

En el siguiente apartado se interpretarán estos valores de cara al filtrado de *outliers*.

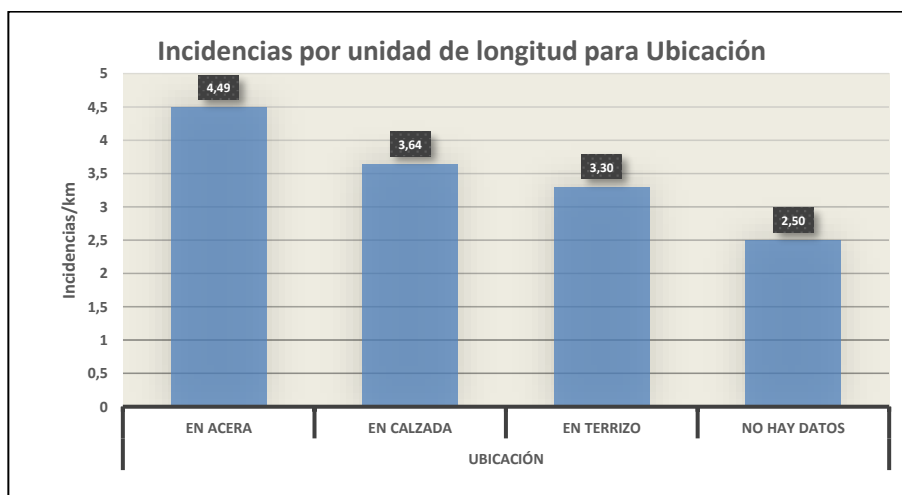


Figura 36. Incidencias por km para factor Ubicación.

2.2.2.4 Distrito

El número de incidencias registradas en el periodo 2005-2018, para la RSA, en el municipio de Sevilla, para el factor Distrito, se recoge en la tabla 24.

	Incidencias	Inc./Long	Long. (m)
SE_BELLAVISTA-LA PALMERA	381	2.7566402	138211.726
SE_CASCO ANTIGUO	576	4.150094449	138792.022
SE_CERRO-AMATE	282	1.466978959	192231.796
SE_ESTE	886	4.175126344	212209.147
SE_LOCAL	231	8.825224677	26174.971
SE_LOS REMEDIOS	185	3.233197705	57218.895
SE_MACARENA	286	2.599735229	110011.203
SE_NERVION	446	4.282542694	104143.737
SE_NORTE	399	1.881142381	212105.157
SE_SAN PABLO-SANTA JUSTA	414	3.288852713	125879.763
SE_SUR	329	2.314009141	142177.485
SE_TRIANA	256	2.88868227	88621.723

Tabla 24. Incidencias acumuladas y por unidad de longitud frente a Distrito.

Estos mismos resultados se representan de forma gráfica, mediante un diagrama de barras en las figuras 37 y 38, respectivamente.

El distrito más afectado por las incidencias es el distrito SE_ESTE, que registra un total de 886 incidencias en el periodo que abarca desde 2005 hasta 2018. Este número de incidencias es equivalente a un 19% de las incidencias totales registradas.

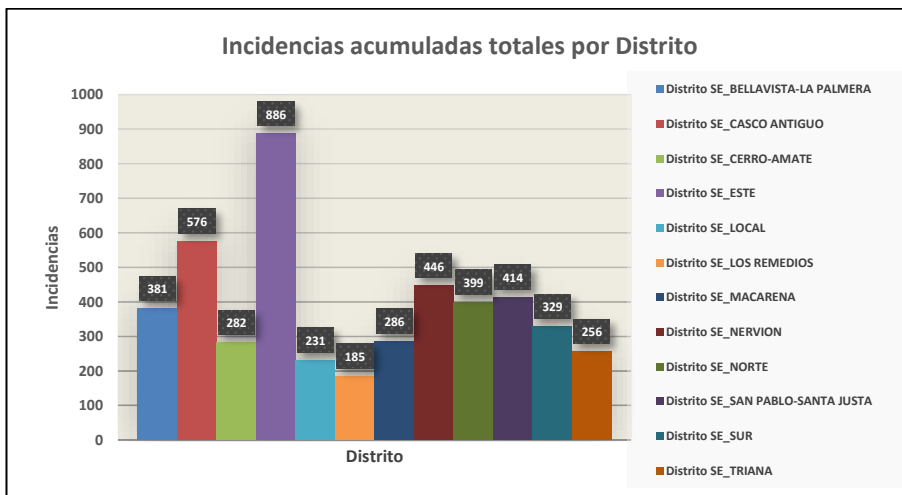


Figura 37. Incidencias por Distrito.

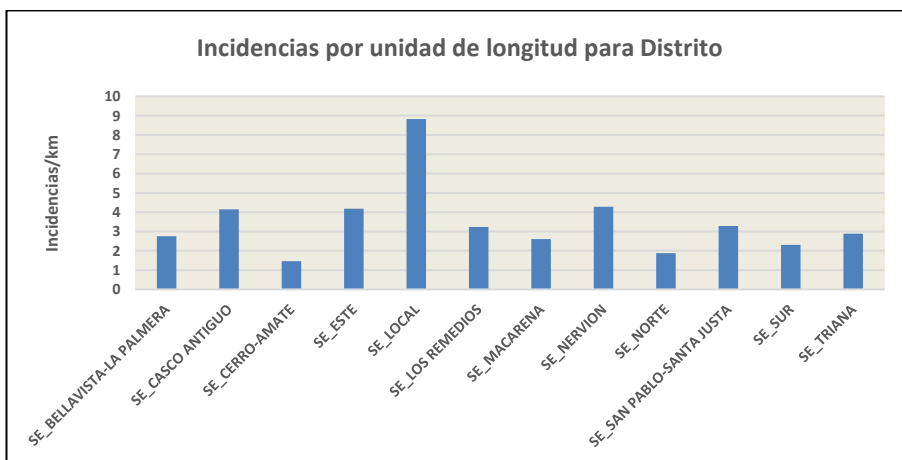


Figura 38. Incidencias por km para factor Distrito.

Por otro lado, se encuentra el distrito SE_CASCO ANTIGUO, con un total de 576 incidencias registradas, equivalente a un 12.33% de las incidencias totales.

En tercer lugar, de los registros que más incidencias registran, se encuentra el distrito de SE_NERVION, con 446 incidencias, iguales a un 10% de las incidencias totales.

En cuanto a los distritos más afectados, destacan SE_LOCAL que registra 8.82 incidencias por unidad de longitud, seguido de SE_NERVION, SE_ESTE, y SE_CASCO ANTIGUO, con 4.28, 4.17 y 4.15 incidencias por unidad de longitud, respectivamente.

El distrito SE_LOCAL es un distrito auxiliar que agrupa a todos aquellos tramos/familias que no tienen un distrito asignado. Representa una longitud muy pequeña de la red, por lo que las incidencias por unidad de longitud asignadas a este distrito presentan un valor muy alto, aun teniendo un número bajo de incidencias acumuladas totales.

2.2.2.5 Año de Instalación

El número de incidencias registradas para el periodo 2005-2018, referentes a la RSA, en el municipio de Sevilla, para el factor Año de Instalación se muestran en la tabla 25. Posteriormente, en las figuras 39 y 40 se representan el número de incidencias acumuladas y por unidad de longitud en el periodo de estudio frente al factor Año de Instalación.

	Año Instalación									
	1910	1950	1960	1965	1968	1970	1971	1972	1974	
Incidencias	23	9	42	13	0	38	0	3	0	
Inc./Long	2.70786834	1.05032529	1.33571383	2.8667885	0	1.75937663	0	1.87313974	0	
Long. (m)	8493.766	8568.774	31443.861	4534.691	1.495	21598.559	1.709	1601.589	36.79	
	1975	1976	1977	1978	1979	1980	1981	1985	1986	1987
	101	0	0	0	5	13	2	15	6	1
	1.73589248	0	0	0	15.5288184	0.94354228	4.59191639	0.82715536	12.4085131	4.15926664
	58183.327	79.499	195.972	2	321.982	13777.867	435.548	18134.441	483.539	240.427
	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997
	0	0	6	10	22	1	0	4	8	2
	0	0	0.2122478	0.3340147	0.85674132	0.51958552	0	0.46791216	1.38775002	0.16592425
	54.356	506.543	28268.844	29938.802	25678.696	1924.611	68.633	8548.613	5764.727	12053.693
	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
	4	1	8	7	3	3	0	0	0	2
	0.38614295	0.09560969	0.81712765	0.60602577	0.0932105	0.24960537	0	0	0	0.16134422
	10358.858	10459.191	9790.392	11550.664	32185.216	12018.972	2547.491	5124.424	38158.458	12395.858
	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
	0	4	1	1	0	1	2	1	3	2
	0	0.41198701	0.12094948	0.12868878	0	1.52462963	0.36787265	0.76061013	1.40224462	0.34100341
	9974.017	9709.044	8267.915	7770.685	6853.91	655.897	5436.664	1314.734	2139.427	5865.044
	2018									
	0									
	0									
	4401.758									

Tabla 25. Incidencias acumuladas e incidencias por unidad de longitud frente a año de Instalación.

La mayoría de incidencias se dan en familias cuyos tramos se instalaron en años previos a 1990. Concretamente, en la década de 1970-1980, es donde se registran los años de instalación de los tramos que más incidencias han sufrido en el periodo 2005-2018. Las familias formadas por los tramos que fueron instalados en 1975 son las que más incidencias registran; del orden de 101 incidencias, seguidos de los tramos que fueron instalados en 1960, y que registran un total de 42 incidencias.

En cuanto a las incidencias por unidad de longitud en función del año de instalación, destacan las fechas de 1979 y 1986, muy superiores a los demás años de instalación, con cifras de 15.52 y 12.4 incidencias por unidad de longitud.

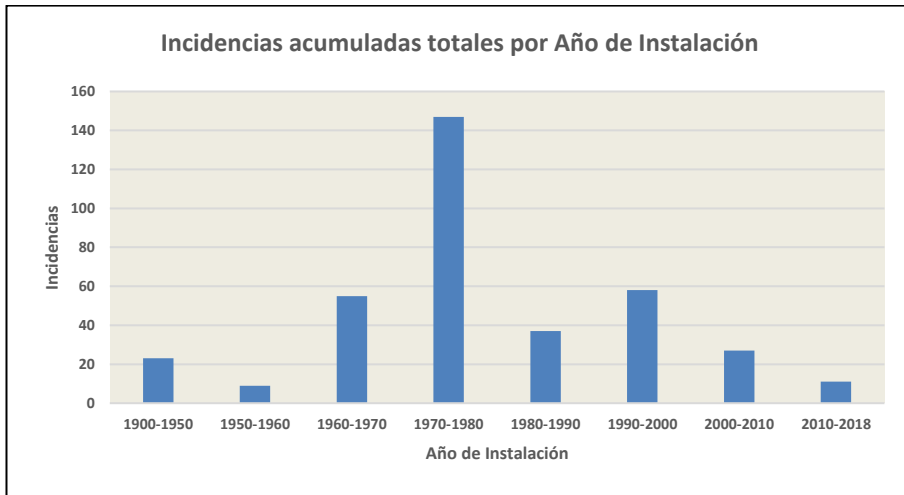


Figura 39. Incidencias acumuladas por Año de Instalación

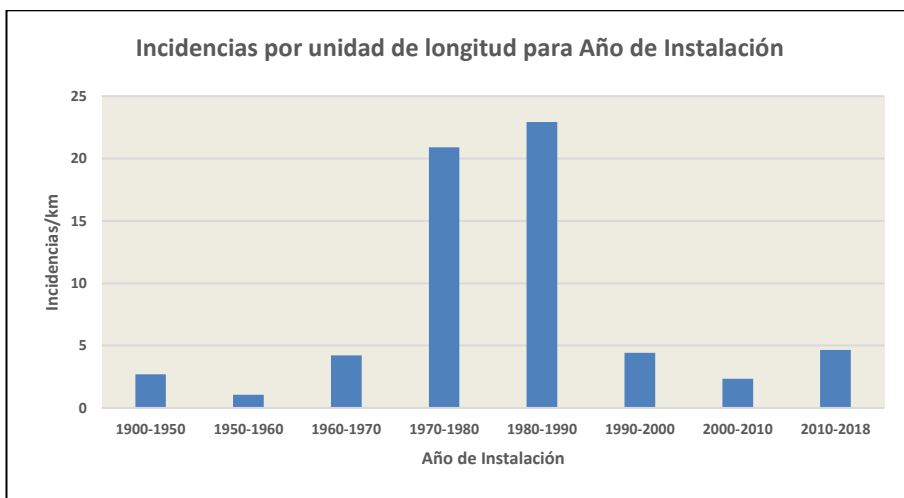


Figura 40. Incidencias por km para factor Año de Instalación.

2.2.2.6 Número de Acometidas

A continuación, se presentan en la figura 41 y 42, la evolución de las incidencias acumuladas y por unidad de longitud, en el periodo 2005-2018, para RSA, en el municipio de Sevilla, en función del factor Número de Acometidas. Resultados en tabla 46 (Anexo).

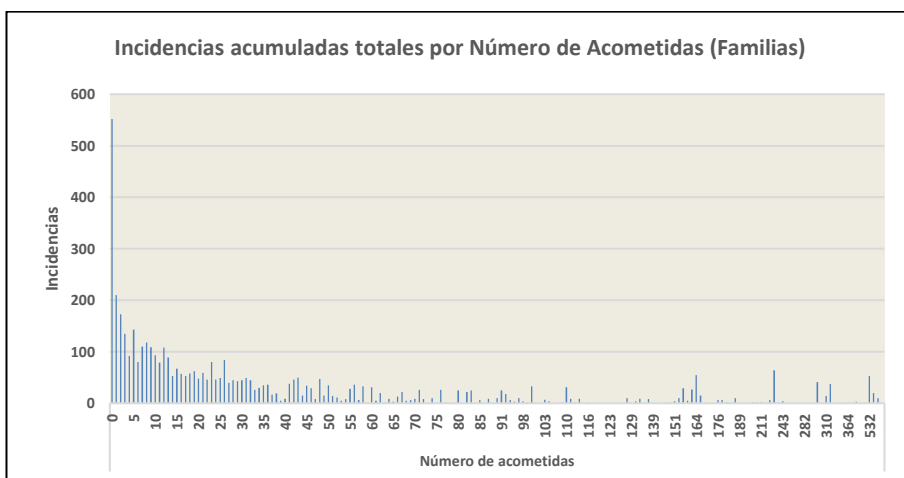


Figura 41. Incidencias acumuladas por Número de Acometidas. Familias.

Conviene recordar que cuando se generaban las familias, el número de acometidas era un factor que sumaba, por lo que el número de acometidas de una familia es la suma del número de acometidas de todos los tramos que la conforman.

La familia que más incidencias presenta engloba un total de 552 incidencias, y está caracterizada por tener 0 acometidas, y una longitud igual a 177292.286 m, respectivamente.

La segunda familia posee 1 acometida, y registra un total de 210 incidencias con una longitud de 68819.38 m.

Existen 50 familias registradas que no presentan incidencias. Sus longitudes varían desde 279 metros, hasta 5007.462 metros, y el rango del número de acometidas de este grupo de familias, engloba desde 59 acometidas hasta 912.

En cuanto a las incidencias por unidad de longitud registradas, se observa que existe una mayor tendencia entre las familias que poseen un número de acometidas entre 150 y 250. Concretamente la familia con un número de acometidas igual a 164, presenta 29 incidencias por unidad de longitud, en una longitud de 1897 metros, seguida de la familia que consta de 231 acometidas, con un total de 26.3 incidencias por unidad de longitud, registradas en 2433.3 metros.

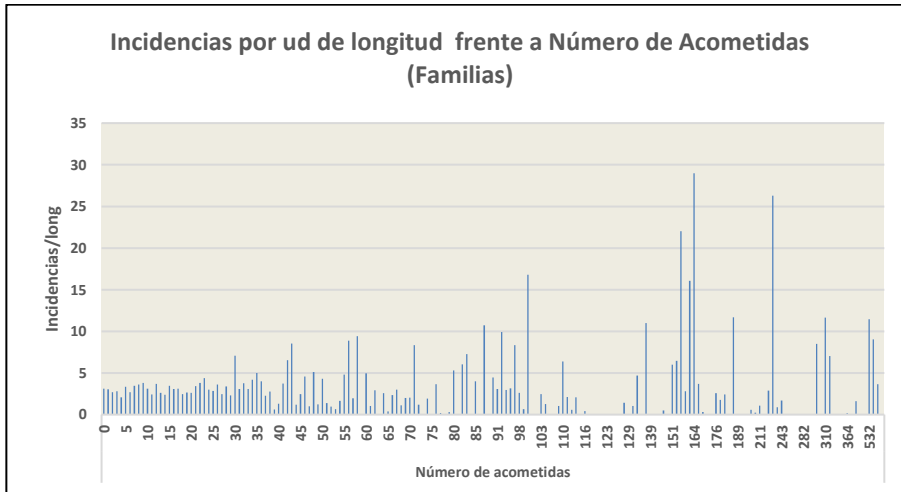


Figura 42. Incidencias por km para factor Número de Acometidas. Familias.

Aun así, debido a que el número de acometidas en la generación de las familias es un factor que suma, es interesante realizar la consulta para los tramos, en este caso, y comprobar a nivel de estudio de tramos, cuáles son los que poseen más tendencia a sufrir una incidencia.

Se presentan estos resultados de forma gráfica en las figuras 43 y 44. Resultados en tabla 47 (Anexo).

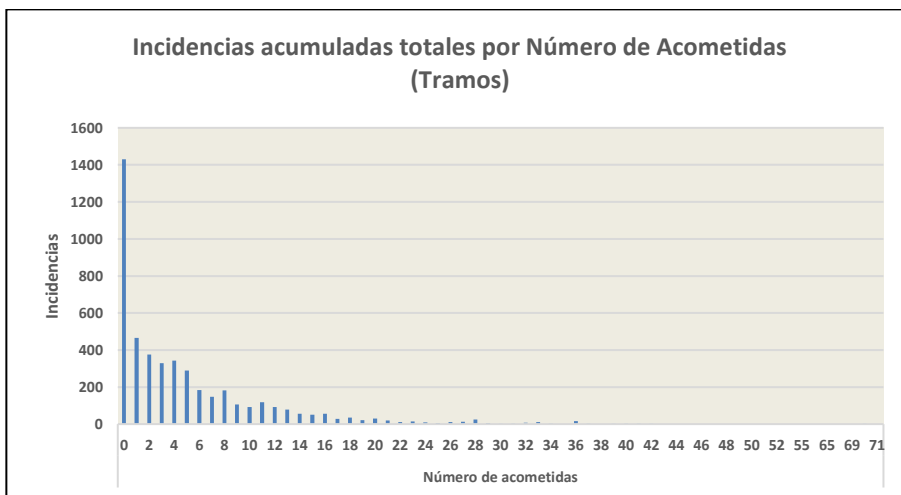


Figura 43. Incidencias acumuladas por Número de acometidas. Tramos.

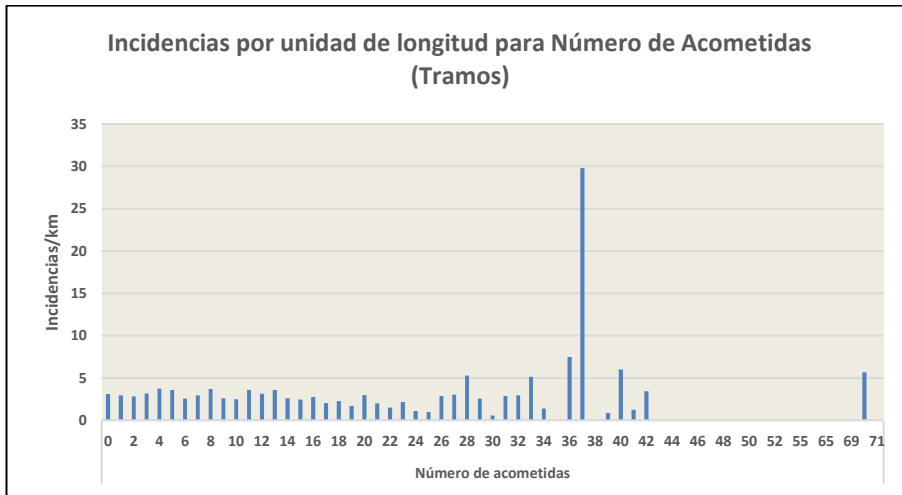


Figura 44. Incidencias por km para factor Número de Acometidas. Tramos.

Tras realizar la consulta para el conjunto de tramos, en lugar del conjunto de familias, se observa que el rango en el que se definen en función del número de acometidas se reduce, y el valor máximo encontrado de acometidas en un tramo es de 71.

En cuanto a las incidencias por unidad de longitud, los valores máximos se dan en tramos que poseen entre 30 y 40 acometidas.

No es necesario relacionar los resultados de familias frente a tramos, ya que como el proceso de filtrado se va a realizar sobre los tramos, y posteriormente se van a generar las familias, esta relación esta automatizada.

2.2.2.7 Presión Media

Dada la gran extensión de registros en función de este parámetro, los resultados se muestran en la tabla 50 (Anexo) de este mismo documento.

Se muestran para el análisis las figuras 45 y 46, que relacionan las incidencias acumuladas con los distintos valores de presión media para las familias generadas en función de los tramos existentes, mediante un diagrama de dispersión.

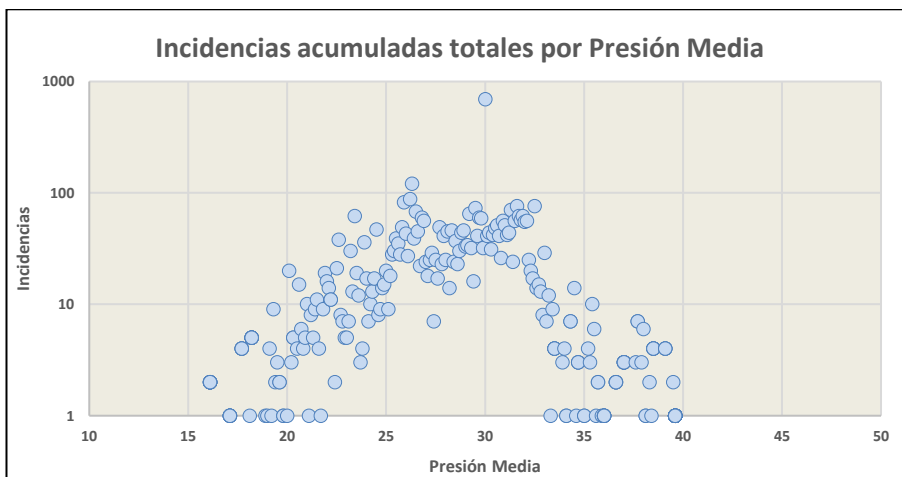


Figura 45. Incidencias acumuladas totales por Presión Media.

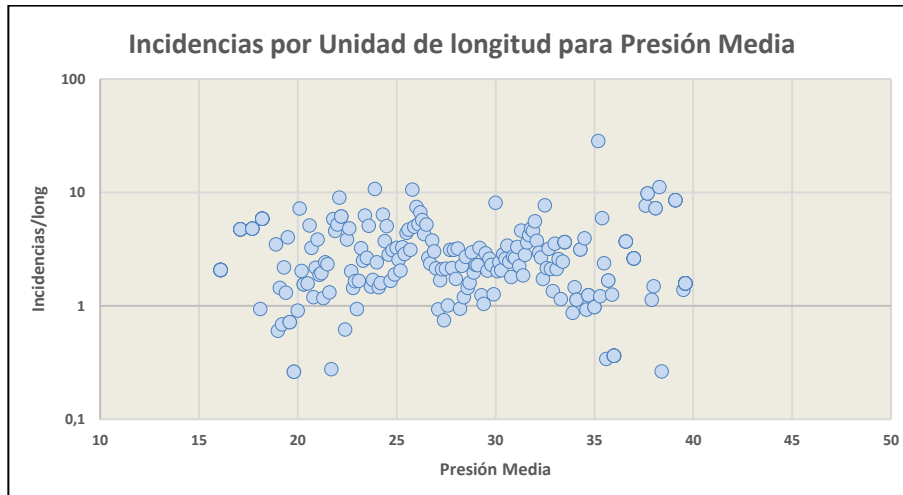


Figura 46. Incidencias por km para factor Presión Media.

La mayoría de las incidencias se dan en familias cuyos tramos están sometidos a una presión media entre 25 y 35 m.c.a.

El mayor número de incidencias, de valor 694, se da en una familia que tiene una presión media de 30 m.c.a, y además consta de una longitud de 85532 m. Se trata de la familia con mayor longitud en esta consulta.

Existen algunas incidencias en familias que están sometidas a presiones medias por encima de 38 m.c.a, pero el número de estas es muy inferior a las mencionadas anteriormente.

En cuanto a las incidencias por km, las familias que tienen más incidencias son las que se encuentran en el rango de presiones entre 20 y 40 m.c.a. Tampoco son notables en este caso, las incidencias por unidad de longitud para las familias que superan una Presión Media por encima de 40 m.c.a.

En este caso, es más representativa la figura que muestra las incidencias por unidad de longitud, puesto que, tanto las incidencias sufridas por una familia, y la longitud de esta misma, se calculan como la suma de todas las incidencias y longitudes de los distintos tramos que la conforman, y, además, la presión media de una familia se calcula como el promedio de las presiones medias de todos los tramos que la forman.

2.2.2.8 Fluctuación de Presión

Se presenta en este apartado, de forma análoga a los apartados anteriores, las consultas realizadas para visualizar la evolución de las incidencias en el periodo 2005-2018 para la RSA, en el Municipio de Sevilla, para el factor Fluctuación de Presión (figuras 47 y 48). Resultados en tabla 53 (Anexo).

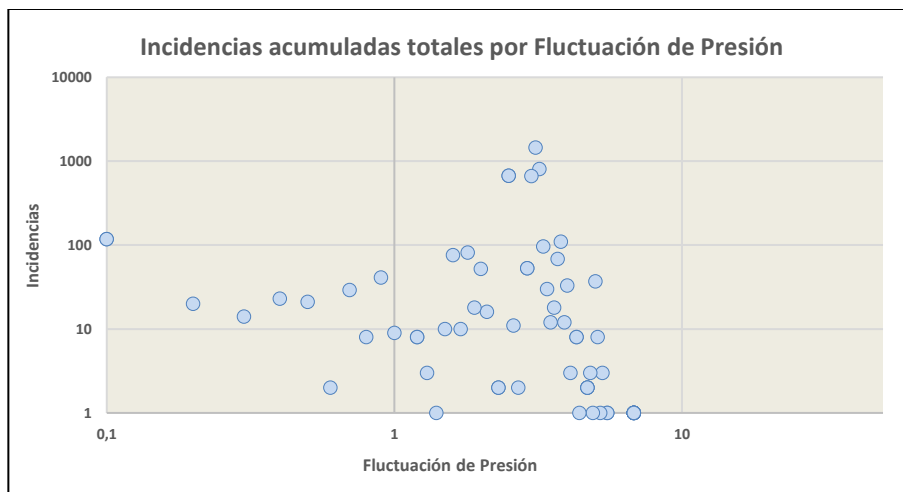


Figura 47. Incidencias acumuladas frente a Fluctuación de Presión.

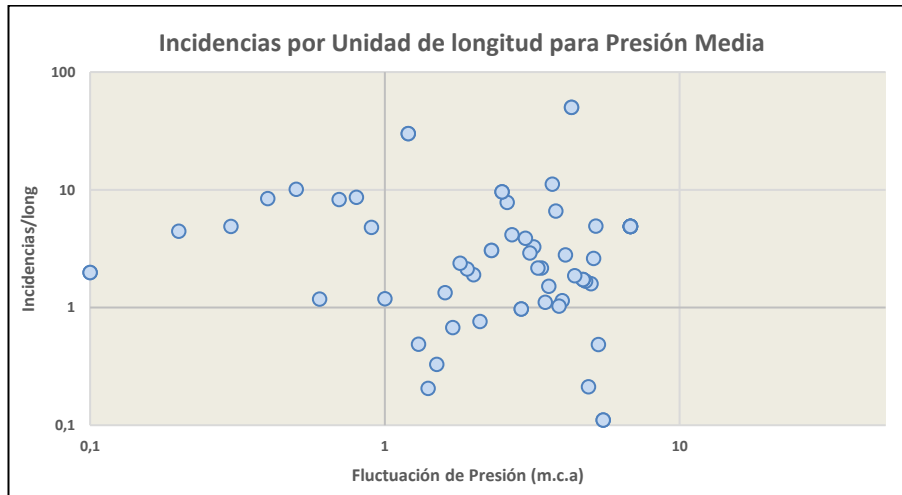


Figura 48. Incidencias por unidad de longitud frente a Fluctuación de Presión.

El mayor número de incidencias están registradas en familias formadas por tramos que están sometidos a unas fluctuaciones de presión entre 2 y 5 m.c.a.

La familia que más incidencias presenta registra un total de 1449 incidencias, con una fluctuación de presión de 3.1 m.c.a.

En cuanto a las incidencias por km, las familias que más recuentan son las que se establecen en un rango de valores de fluctuación de presión, entre 1.5 y 5 m.c.a, análogamente al rango de valores de la fluctuación de presión en el que se dan el mayor número de incidencias acumuladas.

2.2.3 Conclusiones sobre el Análisis de Factores. Filtrado del número de registros.

Tras el análisis de distribución y de incidencias, se van a aplicar una serie de filtros sobre los factores basado en la eliminación de los supuestos *outliers* con el fin de mejorar la calidad de los resultados, basados en la selección de los rangos de valores con mayor concentración de incidencias, y en la relación con la distribución de la longitud en la RSA, en el municipio de Sevilla.

Para eliminar los posibles *outliers*, podrían plantearse dos alternativas:

- La primera consistiría en considerar todos los registros, realizar el ajuste y eliminar posteriormente un cierto porcentaje de los registros, aquellos que están más separados de la curva de ajuste. Dado el número de registros involucrados, esto sólo podría realizarse de forma automatizada.
- El segundo método, consiste en analizar los datos y eliminar los que hayan generado sospechas. Esta opción es la más factible para realizarla de forma manual, y ha sido elegida en este TFG.

Los criterios que se han tomado para generar este filtro se basan en la relación entre las pocas incidencias registradas tanto en nivel global en el período estudiado, como en términos de incidencias por unidad de longitud, y el porcentaje de longitud que representan dentro del marco de la RSA, red para la que se está realizando el análisis.

Finalmente, tras aplicar estos criterios en base a los resultados obtenidos de las consultas en los subapartados 2.1.1 y 2.1.2, se han extraído las siguientes conclusiones sobre los factores:

- **Material.** Se eliminan los registros correspondientes a familias con material PE, PVC y VA, que representan en total el 7.93% de la longitud de la RSA. Ello supone la eliminación de 8364 registros, equivalentes al 7.32%.
- **Diámetro.** Se ha decidido para este ajuste, considerar como outliers los registros definidos por diámetros inferiores a 50 mm e iguales o superiores a 400 mm. Esta reducción supone el filtrado de un 4.9% de los registros.
- **Ubicación.** Frente al histórico de incidencias, se ha comprobado que de los 4 posibles valores que

pueden adoptar las familias en base a este factor el único que no está definido, el cual es la ubicación denominada No hay datos, representa gran parte de la longitud de la RSA. Debido a este motivo, y para no eliminar demasiados registros de cara a la predicción de incidencias, se ha decidido no modificar los registros frente a este factor, y asumir que el indicador No hay datos representa una posible ubicación más, del mismo sentido que En Acera, En Calzada o En Terrizo. Posteriormente se realizará un estudio de influencia de factores y se comprobará el efecto que tiene este factor sobre el cálculo del factor PF. Lo más recomendable sería no incluirlo en el ajuste, aunque en TFG, siguiendo las indicaciones de [2], se va a considerar.

- **Distrito.** En el caso de este factor, tras analizar los resultados se observa una distribución de la longitud de la RSA más equitativa frente a los posibles distritos existentes, que en el caso de otros factores. No se van a realizar modificaciones en los registros para el factor Distrito.
- **Año de Instalación.** Aunque se ha planteado filtrar los registros que cuenten con un año de instalación previo a 1975, dada la gran longitud que representan estos registros, el cribado puede ser demasiado grande y puede no dar los resultados que se esperan. Como conclusión, no se van a aplicar filtros al factor año de Instalación.
- **Número de Acometidas.** Debido a que el factor número de acometidas se define a la hora de generar las familias como un factor que suma, y que los filtros se van a aplicar sobre la hoja de registros de tramos, es conveniente justificar el filtrado en términos de distribución de longitud sobre la distribución de esta en base a los tramos. En términos de tramos, la mayor parte de la longitud de la RSA se define por tramos cuyo número de acometidas es inferior a 40. Los tramos que constan de un número de acometidas superior a 40, representan en términos de longitud el 0.5% de la longitud de la RSA. En términos de incidencias, la mayor parte de las incidencias acumuladas se han registrado en tramos que presentan un bajo número de acometidas. Siguiendo el criterio correspondiente a la longitud, los registros que se definen por un número de incidencias superior a 40 registran el 0.13% de las incidencias totales en la RSA en el período definido.
- **Presión Media.** Dado que el mayor porcentaje de longitud lo representan las familias que están sometidas a una presión media por debajo de 45 m.c.a, y en el análisis del histórico de incidencias se comprobó que, tanto en incidencias acumuladas, como en incidencias por unidad de longitud, las familias conformadas por tramos con presiones superiores a 40 m.c.a presentan un índice muy bajo de incidencias, se ha optado por seleccionar aquellas familias que estén por debajo de 40 m.c.a.
- **Fluctuación de presión.** De forma análoga a factor anterior, la mayoría de la longitud de la RSA está definida por familias que están sometidas a unos valores de la fluctuación de presión entre 0 y 5 m.c.a.

En cuanto al análisis del histórico de incidencias, se observa que la mayoría de incidencias se producen también en tramos con fluctuaciones de presión entre 0 y 5 m.c.a. Del mismo modo ocurre para las incidencias por unidad de longitud frente a este factor.

- **Longitud.** En cuanto al factor longitud, no se ha realizado un análisis específico para el mismo, sino que la justificación realizada para la selección de registros se apoya en [2], donde se concluyó en que los tramos con una longitud muy pequeña penalizan la calidad de las predicciones de la estimación.

Es evidente que, a igualdad del resto de factores, cuanto más larga sea una tubería, más probabilidad existe de que aparezca una incidencia a lo largo de su longitud, pues a más longitud, más superficie hay expuesta a las condiciones o factores causantes de generar una rotura.

En el caso de las familias, debido a que la longitud es un factor que suma, se pueden generar familias con tramos definidos por longitudes pequeñas, lo cual es beneficioso desde el punto de vista de aprovechamiento de datos, y se erradica la problemática derivada de la penalización de la calidad del ajuste por parte de dichos tramos.

En este caso, y como prueba para ver cómo evolucionan los resultados, se van a escoger para el análisis aquellos tramos que tengan una longitud igual o superior a 2 m.

El cambio que se realiza aquí es que, a la hora de generar las familias, dado que el factor longitud

es un factor que suma, estas no agruparán tramos, con sus respectivas características e incidencias acumuladas, que estén dotados de una longitud inferior a la establecida.

Tras aplicar los filtros correspondientes, se pasa de contar en el modelo con un total de 114340 registros, a contar con un modelo seleccionado que cuenta con 60289 registros.

Dentro del grupo de tramos que se encuentran en la zona de Sevilla, y pertenecen a la Red Secundaria de Abastecimiento, las cuales son tanto la zona como la Red objetivo a la hora de ejecutar el modelo de ajuste y predicción, se ha pasado de contar con 50796 registros, a contar con 30117, es decir, se ha realizado una reducción de 20679 registros, equivalente al 40.7% de los datos dotados con estas características.

La bibliografía especializada en ML (Machine Learning) suele considerar entre un 5% y un 10% de *outliers*, por lo que el valor alcanzado aquí es relativamente alto. No obstante, se va a proseguir con los análisis.

Se muestra en la tabla 26, un resumen con las modificaciones realizadas en el fichero de datos para posteriormente generar las hojas de datos que se incluirán en el modelo de ajuste y predicción.

❖ **Resumen de simplificaciones realizadas.**

FACTOR	FILTRADOS REALIZADOS. ELIMINACIÓN DE <i>OUTLIERS</i> .
MATERIAL	Filtrado de los registros designadas por el material VA
	Filtrado de los registros designadas por el material PE
	Filtrado de los registros designados por el material PVC
DIÁMETRO	Filtrado de los registros con diámetros iguales o inferiores a 50 mm.
	Filtrado de los registros con diámetros iguales o superiores a 400 mm.
UBICACIÓN	No se realizan modificaciones para el factor "Ubicación"
DISTRITO	No se realizan modificaciones para el factor "Distrito"
AÑO DE INSTALACIÓN	No se realizan modificaciones para el factor "Año de Instalación"
NÚMERO DE ACOMETIDAS	Filtrado de todos los registros definidos por un número de acometidas superior a 40
PRESIÓN MEDIA	Filtrado de los registros que estén sometidos a una Presión Media por encima de 40 m.c.a.
FLUCTUACIÓN DE LA PRESIÓN	Filtrado de los registros que estén sometidos a una Fluctuación de Presión por encima de 5 m.c.a.
LONGITUD	Filtrado de todos los registros con una longitud inferior a 2 metros

Tabla 26. Procesos de filtrado realizados.

2.3 Análisis del efecto de filtrado de *Outliers*. Comparación de Resultados.

Una vez aplicados los filtros sobre el fichero de datos del que se dispone, y tras haberlas incorporado mediante la macro "Genera Datos" al "Software de Ajuste y Predicción", se presentan a continuación los resultados para los parámetros representativos del Análisis de Bondad realizado para los distintos quinquenios estudiados.

Se han incluido todos los factores para el ajuste. Todos los resultados obtenidos son en función de las **familias**.

En primer lugar, se muestran las tablas comparativas de los valores extraídos para los parámetros representativos de calidad de los resultados (Sensibilidad, Especificidad, Índice de Validez, Valor Predictivo Positivo, Valor

Predictivo Negativo, Índice de Calidad) (tablas 27 y 28). Los resultados de las matrices de confusión se encuentran en las tablas 54-73 (Anexo).

Posteriormente se lleva a cabo la comparación gráfica de la evolución frente a los quinquenios estudiados de los parámetros más importantes para este estudio.

Todos los resultados intermedios referentes a la matriz de confusión de los cuales proceden los valores de los parámetros de calidad se incorporan en el Anexo.

La definición de los distintos parámetros que se van a mostrar a continuación se encuentra en el apartado 1.3.1 Definición de la solución implementada.

❖ Resultados Análisis de Bondad

	Quinquenio 2003-2008.		Quinquenio 2004-2009.	
	SIMPLIFICADOS	GLOBALES	SIMPLIFICADOS	GLOBALES
Sensibilidad %	51.376	44.872	43.781	42.654
Especificidad %	94.347	97.388	96.306	96.51
Índice de Validez %	92.877	95.626	94.65	94.88
Valor predictivo (+) %	24.348	37.367	27.848	27.607
Valor predictivo (-) %	98.208	98.072	98.135	98.18
Índice de Calidad %	37.862	41.12	35.815	35.131

	Quinquenio 2005-2010.		Quinquenio 2006-2011.	
	SIMPLIFICADOS	GLOBALES	SIMPLIFICADOS	GLOBALES
Sensibilidad %	45.714	45.161	48.23	44.444
Especificidad %	96.548	96.714	96.422	96.602
Índice de Validez %	95.152	95.339	94.713	94.852
Valor predictivo (+) %	27.211	27.362	33.131	31.231
Valor predictivo (-) %	98.438	98.47	98.065	98.042
Índice de Calidad %	36.463	36.262	40.681	37.838

	Quinquenio 2007-2012.		Quinquenio 2008-2013.	
	SIMPLIFICADOS	GLOBALES	SIMPLIFICADOS	GLOBALES
Sensibilidad %	43.59	46.078	41.489	38.384
Especificidad %	96.666	96.957	95.167	96.649
Índice de Validez %	95.042	95.468	93.583	94.995
Valor predictivo (+) %	29.21	31.333	20.69	25.083
Valor predictivo (-) %	98.192	98.352	98.166	98.171
Índice de Calidad %	36.4	38.706	31.09	31.734

	Quinquenio 2009-2014.		Quinquenio 2010-2015.	
	SIMPLIFICADOS	GLOBALES	SIMPLIFICADOS	GLOBALES
Sensibilidad %	47.196	38.117	32.877	33.48
Especificidad %	94.399	96.844	95.386	96.694
Índice de Validez %	92.815	94.966	93.238	94.636
Valor predictivo (+) %	22.646	28.523	20.225	25.418
Valor predictivo (-) %	98.094	97.933	97.557	97.737
Índice de Calidad %	34.921	33.32	26.551	29.449

	Quinquenio 2011-2016.		Quinquenio 2012-2017.	
	SIMPLIFICADOS	GLOBALES	SIMPLIFICADOS	GLOBALES
Sensibilidad %	34.894	34.836	38.655	36.992
Especificidad %	96.628	96.716	96.186	96.566
Índice de Validez %	94.352	94.55	94.038	94.464
Valor predictivo (+) %	28.374	27.778	28.221	28.261
Valor predictivo (-) %	97.486	97.615	97.586	97.67
Índice de Calidad %	31.634	31.307	33.438	32.627

Tabla 27. Resumen resultados Datos Simplificados vs Datos Generales.

	MEDIAS	
	SIMPLIFICADOS	GLOBALES
Sensibilidad %	42.7802	40.5018
Especificidad %	95.8055	96.764
Índice de Validez %	94.046	94.9776
Valor predictivo (+) %	26.1904	28.9963
Valor predictivo (-) %	97.9927	98.0242
Índice de Calidad %	34.4855	34.7494

Tabla 28. Promedios Resultados Simplificados vs Globales.

❖ Comparación gráfica de la evolución por quinquenios para los parámetros de calidad

A continuación, se muestran una serie de figuras (figura 49- figura 51) con el objetivo de representar gráficamente la evolución a lo largo de los quinquenios estudiados, de los parámetros estadísticos más relevantes calculados a partir del análisis de bondad. Estos parámetros estadísticos son Sensibilidad, Valor Predictivo Positivo e Índice de Calidad.

Los ejes verticales de las distintas gráficas se han escalado adecuadamente para visualizar mejor la diferencia de valores entre ambas clases de datos.

• Sensibilidad

La sensibilidad se define como la relación o ratio, entre los verdaderos positivos (los tramos que se han estimado que podrían sufrir incidencias y que han coincidido con los reales), y el total de tramos con incidencias reales (verdaderos positivos más falsos negativos).

En la figura 49 se muestra la evolución del parámetro Sensibilidad obtenido a partir del Análisis de Bondad en los distintos quinquenios estudiados para los datos reducidos y los datos originales.

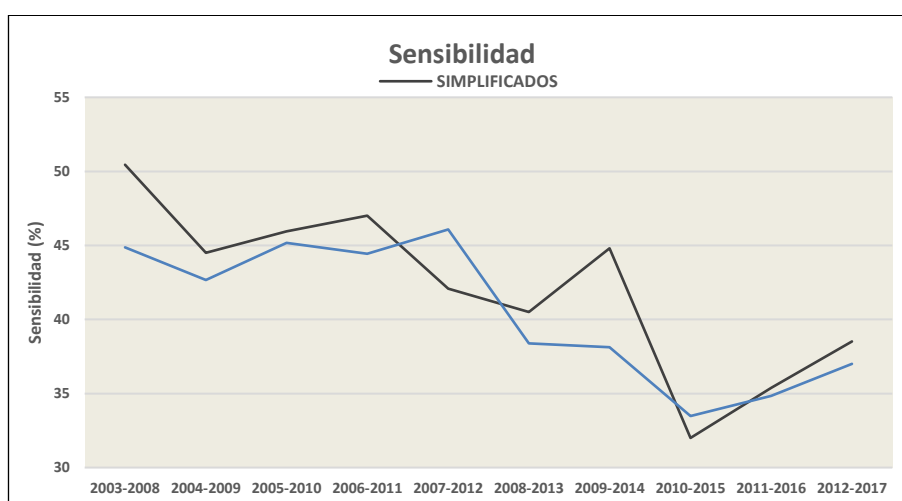


Figura 49. Sensibilidad Datos Simplificados vs Datos Originales.

En general, la sensibilidad obtenida de los resultados de incidencias previstas calculados con los datos filtrados, alcanza un valor más alto en la mayoría de los quinquenios estudiados, excepto en el quinquenio 2007-2012, donde los datos originales alcanzan el máximo en términos de la sensibilidad.

En el quinquenio 2010-2015 se aprecia una caída muy considerable de la sensibilidad en relación con los datos filtrados, lo cual hace que la sensibilidad de los datos originales supere el valor de esta, por muy poco.

Se observa que la sensibilidad en ambos casos va disminuyendo hasta que en el quinquenio 2010-2015 alcanza el mínimo y vuelve a aumentar.

• Valor Predictivo Positivo

El valor predictivo positivo es uno de los parámetros más importantes para el interés de este problema en la comprobación de calidad de las estimaciones. Se define como la relación o el cociente entre el número de tramos con incidencias estimado que han coincidido con las incidencias reales (VP), y el número total de incidencias estimado (VP+FP). Se muestran los resultados en la figura 50.

Se puede observar que la evolución de este parámetro sigue una trayectoria parecida en los dos casos, adoptando para el caso del uso de datos totales valores un poco mayores que en el caso de los datos filtrados.

No obstante, esta diferencia no es notable, pues se está hablando de variaciones de entorno al 5% entre un caso y el otro.

En el primer quinquenio es donde se da la diferencia más notable, que es de 24.349% en el caso de los datos filtrados, frente a 37.367% en el caso del conjunto de datos totales, observando una variación del 13%.

En los siguientes periodos la diferencia disminuye hasta casi igualar los valores entre ambos grupos, y a partir del quinquenio 2009-2014, vuelve a caer el valor en el caso de los datos filtrados, pero con diferencias muy poco apreciables.

Existen puntos donde el valor predictivo positivo en el caso del grupo de datos sometido a filtrado supera al conjunto de datos global.

En cualquier caso, se observa que no se produce una mejora de la calidad de los resultados, sino que la calidad se mantiene para constante (con diferencias muy poco apreciables, menos en el caso del primer quinquenio) para ambos conjuntos de datos (originales y filtrados).

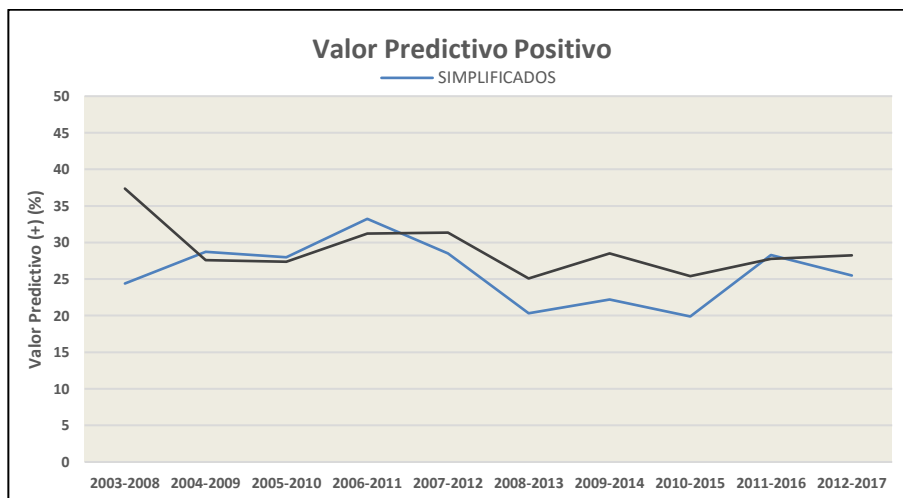


Figura 50. Valor Predictivo Positivo Datos Simplificados vs Datos Globales.

- **Índice de Calidad**

El índice de calidad se define por la semisuma de la sensibilidad y el valor predictivo positivo, los cuales son los dos parámetros estadísticos más representativos a la hora de realizar el estudio de calidad de resultados del modelo predictivo en el análisis de bondad. Se trata de un parámetro no estándar, sino que tiene su origen en el desarrollo de la investigación realizada por GERM [2], con el objetivo de tener una visión general de la calidad de los resultados.

Como se puede apreciar en la figura 51, ambos conjuntos de datos mantienen una evolución de dicho factor muy parecida, adoptando en algunos casos el grupo de datos filtrados valores del índice de calidad superiores al grupo de datos brutos (sin tratar), y viceversa.

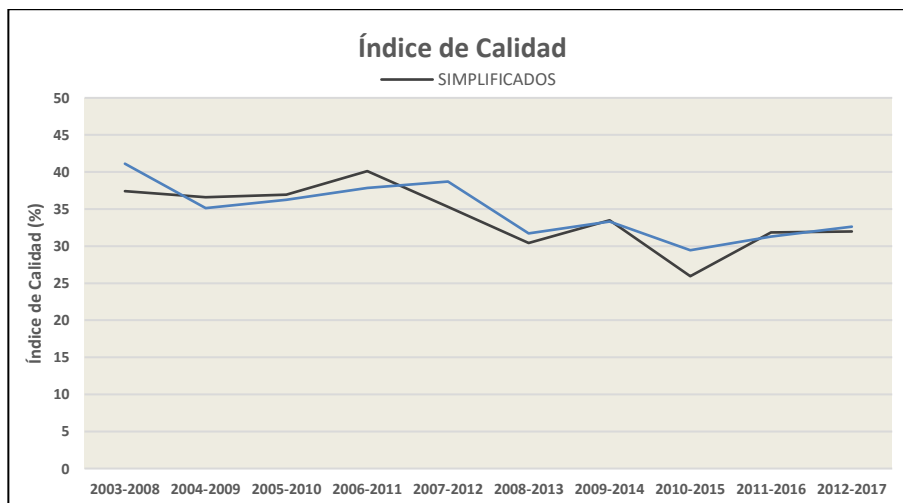


Figura 51. Índice de calidad para datos Simplificados vs Datos Globales.

Las variaciones o diferencias entre un grupo y otro no son notables, siendo las más destaca la obtenida en el primer quinquenio de estudio (2003-2008), donde se muestra una variación de un 3.26%, manteniendo el grupo de datos globales un valor de IC de 41.12% y 37.86% en el caso de los datos simplificados, respectivamente.

Aun así, ambos resultados son bastante satisfactorios, pues presentan un alto porcentaje de aciertos.

Es curioso observar, como en el caso del Valor Predictivo Positivo, que aun habiendo reducido el conjunto de registros en un 40%, los resultados en base al IC (Índice de Calidad), son muy semejantes.

❖ Comparación de Valores Medios

Se muestra a continuación una figura representativa (figura 52) de la comparación de los valores medios para los distintos parámetros estadísticos obtenidos del análisis de bondad mediante la herramienta Matriz de Confusión, para observar cómo es la variación entre el caso del uso de datos totales, y el caso de los datos simplificados.

Estos resultados provienen de realizar el promedio de los valores obtenidos para cada parámetro en cada quinquenio.

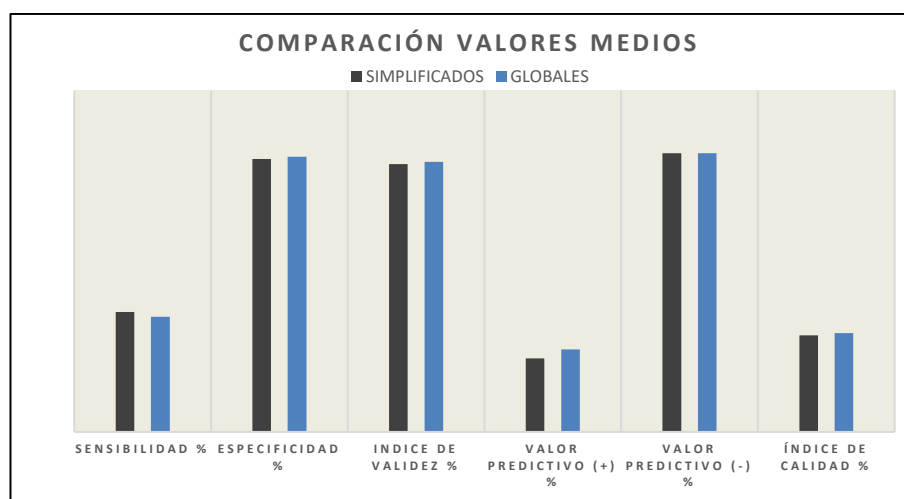


Figura 52. Comparación Valores Medios Datos Simplificados vs Datos Globales.

Se puede observar que, en la mayoría de los casos, la diferencia es casi inapreciable.

En el caso de la sensibilidad está justificado que el grupo de datos filtrados aporta una mejor resolución a la predicción que el uso de datos globales, siendo la variación de 1.62% favorable al primero.

En el caso de la especificidad, aumenta un poco para el caso de los datos globales, dando como resultado una diferencia de 0.814%.

La variación en función del Índice de Validez es casi inapreciable, con una diferencia que no llega al 1% entre el grupo de datos simplificados y el grupo de datos globales.

En el valor predictivo positivo es donde más notable se hace esta diferencia, pero, aun así, no llega al 5%, siendo el valor de la diferencia del 3%.

En el caso del valor predictivo negativo, el grupo de datos simplificados obtiene un valor mayor frente al grupo de datos globales, siendo la diferencia de entorno al 0.02%. Prácticamente nula.

Por último, el IC (Índice de Calidad) presenta una diferencia del 0.7% favorable al grupo de datos globales.

Como introducción a las conclusiones que se realizarán en el capítulo 4 de este documento, se concluye con que no se ha alcanzado la mejora de la calidad de los resultados.

Por un lado, el análisis de los resultados muestra que parece ser que los datos filtrados no han influido en el modelo de regresión, aunque, desde el punto de vista estadístico, no se deberían de realizar los

filtros con el objetivo de la mejora de calidad, o al menos, no con el procedimiento que se ha realizado. Finalmente se concluye con que no merece el filtrado de *outliers*, del modo en que se ha realizado. Una propuesta futura sería eliminar aquellos registros más alejados del ajuste.

3 TRATAMIENTO DE DATOS Y CONTRASTE DE RESULTADOS

En este capítulo, se aplicarán una serie de tratamientos sobre ciertos factores con el fin de mejorar los resultados en la regresión.

Los tratamientos que se realizan son dos:

- En primer lugar, se analizarán los factores cuantitativos (Diámetro, Número de Acometidas, Presión Media, etc.) (véase 1.3.2 Factores considerados), con el fin de aplicar el tratamiento de normalización a los datos de entrada o *inputs*.
- En segundo lugar, se realizarán algunas pruebas mediante periodos elegidos al azar con la finalidad de evidenciar la relación existente entre el factor PF, y la Probabilidad de Fallo real de los tramos/familias.

El capítulo se estructura en dos apartados generales:

- Un primer apartado donde se aplica el procedimiento de normalización, que a su vez se estructura en cinco subapartados: el primer subapartado dedicado a la justificación de por qué se aplica el procedimiento de normalización; un segundo subapartado que se centra en explicar en qué consiste el procedimiento, un tercer apartado que tiene como objetivo el estudio de qué factores son susceptibles de someterse al procedimiento de normalización; el cuarto apartado se centra en mostrar los resultados obtenidos en el análisis de bondad mediante quinquenios (fase de comprobación) con los datos normalizados frente a los datos originales; y un quinto capítulo donde se realiza una interpretación de los coeficientes de ajuste que multiplican a los factores que intervienen en el modelo de regresión, con la finalidad de estudio de pesos o influencia de los factores.
- Un segundo apartado donde se realiza una comparación entre el índice que asigna el modelo de ajuste y predicción, PF, a cada tramo o familia, y la Probabilidad de Fallo real de las familias, calculada mediante el histórico de incidencias disponible para tres períodos aleatorios. La idea principal de este apartado es aportar evidencias sobre la relación entre el parámetro PF y la Probabilidad de Fallo real de las familias.

3.1 Análisis de Factores Cuantitativos.

3.1.1 Justificación del proceso de Normalización.

La gran ventaja que presenta la aplicación de un tratamiento de normalización al conjunto de datos que componen el volumen de registros con el que se está trabajando, es la capacidad de poder comparar de manera adimensional factores, que, en principio, tienen una naturaleza muy distinta.

En el modelo de regresión intervienen factores como el Diámetro, con escala en mm, el número de acometidas, con escala en acometidas, las distintas presiones, con escala en m.c.a, y todos los factores cuantitativos que se quieran añadir a posteriori de este estudio. Esto hace que los coeficientes que se obtienen directamente de la fase de ajuste, no se puedan comparar entre sí (de forma objetiva), dado que las escalas son muy variadas. Además, impiden también analizar la evolución de los coeficientes que acompañan a los factores cualitativos, ya que se trata de una relación lineal, donde los factores cuantitativos intervienen en el mismo nivel que los cualitativos.

Normalizar los datos permite utilizar factores con una escala común, donde los coeficientes de ajuste que acompañan a cada factor ya sí pueden ser comparables entre sí, dado que todos los factores que

intervienen en el modelo son adimensionales, mientras que los coeficientes de la regresión tienen unidades de incidencias esperadas.

Existe una gran variedad de transformaciones dentro del campo del tratamiento de datos en análisis de investigación estadística.

De entre las distintas posibilidades de transformación de datos mediante normalización, se pueden encontrar dos tipos principales de normalización según CEPAL (Comisión Económica para América Latina y el Caribe) (fuente: <https://biblioguias.cepal.org/c.php?g=495473&p=4398013>):

- **Normalización estadística:** que consiste principalmente en ajustar los valores medidos en diferentes escalas, a una escala común, adimensional, donde se puedan comparar directamente los pesos de los distintos factores entre sí, y se pueda estudiar el efecto que cada uno tiene sobre la aparición de incidencias en un determinado tramo.
- **Normalización de base de datos:** que tiene como principal objetivo eliminar mediante reglas de diseño relacionales posibles valores duplicados e incongruencias con el objetivo de mejorar la calidad de los datos al entrar. El proceso de filtrado realizado en el Capítulo 2, puede ser considerado un procedimiento de este tipo.

Según [5] existen a su vez varias normalizaciones estadísticas para la mejora de calidad de datos. Entre otras, se pueden encontrar las siguientes:

- **Puntuación estándar o tipificada:** tiene como objetivo normalizar los errores cuando los parámetros de población son conocidos.
- **T de student:** se trata de una normalización residual donde, en este caso, los parámetros de población son desconocidos.
- **Studentized Residual:** se trata de una normalización residual cuando los parámetros son estimados, especialmente a través de puntos de diferentes datos en análisis de regresión.

Existen también, otros como el momento estandarizado, el coeficiente de variación, la característica de Scaling, etc.

Debido a que en este análisis se cuenta con los datos reales de población, se va a implementar la normalización estadística “Puntuación Estándar”.

3.1.2 Proceso de Normalización.

El proceso de normalización se puede realizar de forma manual, factor a factor, o implementándolo en un software concreto para el tratamiento de los distintos registros que conforman el sistema de datos de entrada.

En este caso, los factores no son tan numerosos y se puede aplicar perfectamente la primera opción, donde el procedimiento consiste básicamente en calcular la media del conjunto de valores que constituyen cada factor cuantitativo, y que están presentes en la definición de los registros de entrada. Por otro lado, hay que calcular la desviación típica, la cual es otro parámetro de entrada en el proceso de normalización.

Una vez calculados ambos parámetros, se procede a normalizar, tomando la expresión de la puntuación estándar que se muestra a continuación:

$$z_i = \frac{x_i - \mu}{\sigma}$$

Siendo μ la media, σ la desviación estándar, y x_i los distintos valores sin normalizar que adopta cada factor a lo largo de los distintos registros que conforman el grupo de datos de entrada.

La media se calcula como el promedio de los valores pertenecientes a un factor cuantitativo en concreto, que definen los registros.

$$\mu = \frac{\sum_{i=1}^N x_i}{N}$$

La desviación estándar se calcula mediante la siguiente expresión:

$$\sigma = \sqrt{\frac{\sum |x - \mu|^2}{N}}$$

Donde N es el número de valores que se incluyen en el cálculo.

Con el fin de no realizar estos cálculos de forma manual debido a la facilidad de introducir errores, se ha optado por generar dichos valores en una hoja de Excel, mediante los comandos “PROMEDIO”, y “DESVEST.M”.

Una vez calculados dichos parámetros, se genera una columna análoga a la de la muestra objeto de tratamiento, generando los valores normalizados para cada parámetro.

Tras obtener los valores normalizados, se sustituyen en la hoja original de datos que se obtuvo desde las distintas bases de datos de EMASESA, y se generan mediante la herramienta “Genera Datos” los ficheros correspondientes tanto en el software para análisis de datos como en el modelo de ajuste y predicción.

Una vez introducidos dichos valores en los modelos, se procede a realizar el análisis de bondad para comprobar si gracias a realizar este tratamiento existe una mejora en la calidad de los resultados.

Los factores que se consideran en el modelo de ajuste y predicción son todos los factores que caracterizan los registros de datos y que se han utilizado hasta el momento, excluyendo únicamente el factor longitud, dado que como se ha comentado anteriormente, no puede someterse al proceso de normalización.

Los intervalos de PF se siguen manteniendo, siendo $PF \geq 0.75$ alto, $PF > 0.5$ medio, y $PF \leq 0.5$ bajo.

El nivel de confianza para las predicciones de incidencias totales se mantiene en 97.5%.

Se utilizan todos los datos salvo los registros correspondientes a familias con longitud inferior a 3m (véase [2]). A su vez, se acotarán los registros utilizados a la RSA y al municipio de Sevilla.

3.1.3 Estudio de Factores Susceptibles de Normalización.

Sólo pueden ser susceptibles de normalización aquellos factores englobados en el grupo de factores cuantitativos.

Los factores cuantitativos que se definen en el grupo de datos extraídos desde las correspondientes bases de datos de EMASESA son los siguientes:

- Diámetro
- Edad
- Longitud.
- Número de acometidas
- Presión mínima
- Presión media
- Presión máxima

No obstante, y en este caso particular, la condición de susceptibilidad frente a normalización no depende sólo de que el factor pueda ser medible en términos numéricos, sino que existen ciertos factores cuantitativos que no pueden normalizarse dentro de este procedimiento.

- **Diámetro.** El diámetro es un factor susceptible de someterse a un tratamiento de normalización. Su rango de valores abarca desde 20 mm hasta 1700 mm. Tras someterse al proceso de normalización,

el rango obtenido para este factor abarca desde -1 hasta 12.56.

- **Edad.** La edad no es un factor como tal que pueda identificarse directamente en los registros obtenidos a través de la base de datos de EMASESA. El factor directamente relacionado con este parámetro es el factor Año de Instalación. La edad se calcula como la diferencia entre el año sobre el que se realiza el ajuste o la predicción de incidencias en un determinado tramo, conjunto de tramos, familia, o conjunto de familias, y el año de instalación del objeto en cuestión. Por eso no se puede normalizar de la misma forma que se ha hecho con el resto de factores cuantitativos.
- **Longitud.** A la longitud no puede aplicársele el tratamiento de normalización diseñado debido a que su valor interviene en operaciones internas del software utilizado, y por lo tanto no se debe modificar.
- **Número de acometidas.** De forma natural, el factor conforma los posibles valores que puede adoptar en un rango que va desde 0 acometidas hasta 71 acometidas, en términos del número de acometidas que pueden ser encontradas en tramos.

Tras someterse al proceso de normalización, los valores que adopta este parámetro se engloban en el rango que va desde -0.39 hasta 15.95.

- **Presión mínima, media y máxima.** Este grupo de factores, de la misma naturaleza, también pueden someterse a un procedimiento de normalización debido a cuestiones análogas a las de los factores cuantitativos susceptibles de someterse a normalización descritos anteriormente.

Los rangos naturales que pueden adoptar estos valores son: en el caso de la Presión Mínima el rango engloba desde 0.19 m.c.a hasta 365.5 m.c.a, en el caso de la Presión media, este rango va desde 0.19 m.c.a hasta 365.6 m.c.a, y finalmente en el caso de la Presión Máxima, este rango agrupa desde 0.2 m.c.a hasta 365.8 m.c.am respectivamente.

Una vez normalizados, los rangos pasan a ser: desde -3.216 hasta 38.264 en el caso de la presión mínima, desde -3.477 hasta 39.139 en el caso de la presión media, y desde -3.6 hasta 39.6 en el caso de la presión máxima.

En la tabla 29 se muestran los valores tanto de la media como de la desviación estándar para cada parámetro susceptible de someterse al proceso de normalización por puntuación estándar o tipificada. Además, en las tablas 30 y 31 se muestra un ejemplo de la estructura de los valores de los factores originales, y una vez sometidos al proceso de normalización.

Se muestra también en la figura 53 un ejemplo de la estructura final de los registros con los factores normalizados una vez realizada la inclusión de los mismos en el modelo de ajuste y predicción.

	PROMEDIO	DESV. ESTÁNDAR
Diametro nominal	144.11809	123.8151637
Número de acometidas	1.70522302	4.343140987
Presión mínima (mca)	28.513542	8.806855351
Presión media (mca)	30.0069044	8.574357482
Presión máxima (mca)	30.9734535	8.529195548

Tabla 29. Valores De promedio y desviación estándar para Normalización.

Diametro nominal (mm)	Número de acometidas	Presión mínima (mca)	Presión media (mca)	Presión máxima (mca)
150	0	0.93	1.04	1.09
200	0	0.98	24.83	33.81
150	0	1.22	25.29	50.2
150	10	1.22	25.29	50.2
150	0	1.24	25.31	50.22
150	0	1.49	25.56	50.47
...

Tabla 30. Datos Originales.

Diametro nominal	Número de acometidas	Presión mínima	Presión media	Presión máxima(mca)
0.047505569	-0.392624376	-3.132053486	-3.378317787	-3.503666118
0.451333327	-0.392624376	-3.126376091	-0.603765868	0.33256906
0.047505569	-0.392624376	-3.099124593	-0.550117534	2.25420398
0.047505569	1.909856715	-3.099124593	-0.550117534	2.25420398
0.047505569	-0.392624376	-3.096853635	-0.547784997	2.256548867
0.047505569	-0.392624376	-3.068466658	-0.518628294	2.285859955
...

Tabla 31. Datos Normalizados.

Figura 53. Estructura de datos normalizados en Modelo de Ajuste y Predicción.

3.1.4 Resultados por Quinquenios. Contraste de Resultados.

En este subapartado se procede a mostrar y comparar los resultados obtenidos en el análisis de bondad realizado por quinquenios tanto para los datos originales como para los datos tras implementar tratamiento de normalización para los factores cuantitativos comentados anteriormente. Para obtener dichos resultados se han considerado las familias que tienen una longitud igual o superior a 3 m, que se encuentran en el municipio de Sevilla, y que pertenecen a la RSA.

La tabla 32 representa un resumen de resultados y comparativa para los distintos análisis de bondad realizados por quinquenios mediante el uso de datos con factores cuantitativos normalizados frente a factores sin normalizar (resultados matrices de confusión tablas 74-93 (Anexo)).

	QUINQUENIO 2003-2008		QUINQUENIO 2004-2009	
	NORMALIZADOS	SIN NORMALIZAR	NORMALIZADOS	SIN NORMALIZAR
Sensibilidad %	40.249	42.667	36.123	40.284
Especificidad %	97.257	96.592	97.091	96.436
Índice de Validez %	95.502	94.852	95.323	94.737
Valor predictivo (+) %	31.803	29.448	27.063	26.074
Valor predictivo (-) %	98.085	98.059	98.072	98.104
Índice de Calidad %	36.026	36.058	31.593	33.179

	QUINQUENIO 2005-2010		QUINQUENIO 2006-2011	
	NORMALIZADOS	SIN NORMALIZAR	NORMALIZADOS	SIN NORMALIZAR
Sensibilidad %	39.13	43.011	42.57	45.299
Especificidad %	97.296	96.655	97.281	96.632
Índice de Validez %	95.757	95.224	95.54	94.909
Valor predictivo (+) %	28.223	26.059	33.974	31.832
Valor predictivo (-) %	98.328	98.41	98.097	98.072
Índice de Calidad %	33.677	34.535	38.272	38.566

	QUINQUENIO 2007-2012		QUINQUENIO 2008-2013	
	NORMALIZADOS	SIN NORMALIZAR	NORMALIZADOS	SIN NORMALIZAR
Sensibilidad %	43.805	48.529	36.199	36.364
Especificidad %	97.579	97.031	97.37	96.59
Índice de Validez %	96.026	95.612	95.642	94.88
Valor predictivo (+) %	34.982	33	28.571	23.762
Valor predictivo (-) %	98.316	98.426	98.131	98.111
Índice de Calidad %	39.394	40.765	32.385	30.063

	QUINQUENIO 2009-2014		QUINQUENIO 2010-2015	
	NORMALIZADOS	SIN NORMALIZAR	NORMALIZADOS	SIN NORMALIZAR
Sensibilidad %	33.471	37.668	30.165	33.04
Especificidad %	97.468	96.83	97.376	96.68
Índice de Validez %	95.489	94.938	95.297	94.608
Valor predictivo (+) %	29.67	28.188	26.838	25.084
Valor predictivo (-) %	97.868	97.918	97.762	97.723
Índice de Calidad %	31.571	32.928	28.502	29.062

	QUINQUENIO 2011-2016		QUINQUENIO 2012-2017	
	NORMALIZADOS	SIN NORMALIZAR	NORMALIZADOS	SIN NORMALIZAR
Sensibilidad %	30.038	34.426	34.483	37.805
Especificidad %	97.421	96.701	97.329	96.596
Índice de Validez %	95.157	94.522	95.233	94.522
Valor predictivo (+) %	28.832	27.451	30.822	28.882
Valor predictivo (-) %	97.563	97.6	97.73	97.7
Índice de Calidad %	29.435	30.939	32.653	33.344

Tabla 32. Resumen Resultados por quinquenios Datos Normalizados vs Sin Normalizar.

La tabla 33 muestra los resultados promedio de los distintos parámetros estadísticos resultado de los análisis de bondad para los quinquenios estudiados.

	VALORES MEDIOS	
	NORMALIZADOS	SIN NORMALIZAR
Sensibilidad %	36.6233	39.9093
Especificidad %	97.3468	96.6743
Índice de Validez %	95.4966	94.8804
Valor predictivo (+) %	30.0778	27.978
Valor predictivo (-) %	97.9952	98.0123
Índice de Calidad %	33.3508	33.9439

Tabla 33. Promedio Resultados por quinquenios Datos Normalizados vs Sin Normalizar.

❖ Figuras Comparativas de Resultados

La sensibilidad proporciona el porcentaje de incidencias estimadas que coinciden con las reales frente al total de incidencias reales que se han producido en el año en concreto usado para la predicción, mientras que el valor predictivo positivo proporciona un porcentaje de incidencias acertadas, es decir, incidencias estimadas que coinciden con las reales, frente al total de incidencias estimadas, por lo que el valor predictivo positivo proporciona una mayor fiabilidad frente a la calidad del método.

○ Sensibilidad

La figura 54 representa la evolución del factor Sensibilidad a lo largo de los quinquenios analizados para los datos normalizados frente a los datos sin normalizar.

La evolución a lo largo de los quinquenios del factor sensibilidad sigue una dirección análoga en ambos casos.

Se aprecia que, en este caso, los datos que no han sido sometidos a tratamiento de normalización obtienen mejores resultados, aunque la diferencia es muy pequeña entre ambos.

Los valores medios para los dos casos de la sensibilidad a lo largo de los diez quinquenios estudiados son: 36.6% de sensibilidad media para el caso de los datos normalizados, y de 39.9% para el caso de los datos sin normalizar.

En el quinquenio 5, que abarca desde el año 2007 hasta el 2012, es donde se da la diferencia más grande entre valores de sensibilidad en la comparación, existiendo una diferencia del 4.724% a favor de los resultados con los datos sin normalizar.

El hecho de que el valor de la sensibilidad sea más alto en el caso de los resultados del análisis de bondad mediante los datos sin normalizar, no quiere decir que el uso de ese grupo de datos sea más eficaz, o proporcione mejores resultados.

Existen cinco parámetros más que es preciso analizar para poder hallar una conclusión sostenible acerca de la fiabilidad del método de normalización.

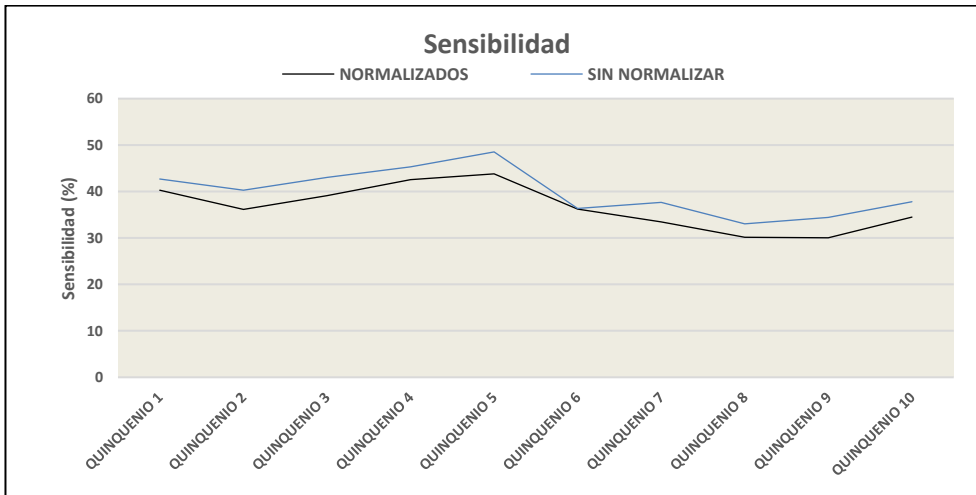


Figura 54. Sensibilidad para Datos normalizados vs Datos sin normalizar.

○ **Valor Predictivo (+)**

El parámetro Valor Predictivo Positivo constituye uno de los valores más importantes en referencia al interés de este problema, a la hora de comparar los resultados de las predicciones mediante el análisis de bondad.

Este valor proporciona la relación entre los verdaderos positivos, es decir, aquellos tramos con incidencias estimados y que coinciden con los tramos con incidencias reales, y el total de positivos estimados, es decir, los verdaderos positivos, y los falsos positivos, aquellos que se han estimado como tramos con incidencias pero que realmente no lo eran. Además, da pie a justificar la mejora del proceso de predicción dado que cuanto mayor sea, más se acercan los resultados estimados a los resultados reales en el análisis de bondad.

Se muestra la evolución para este parámetro en función de los datos estudiados en la figura 55.

En la figura se puede observar que los resultados para el Valor Predictivo Positivo son en todo momento mayores para el caso de la ejecución del modelo de predicción con los datos sometidos a tratamiento que en el caso de la predicción con los datos originales.

Sorprendentemente, en el quinquenio 6 se produce un incremento del índice de calidad para el grupo de datos normalizados de casi un 5%, incremento muy considerable.

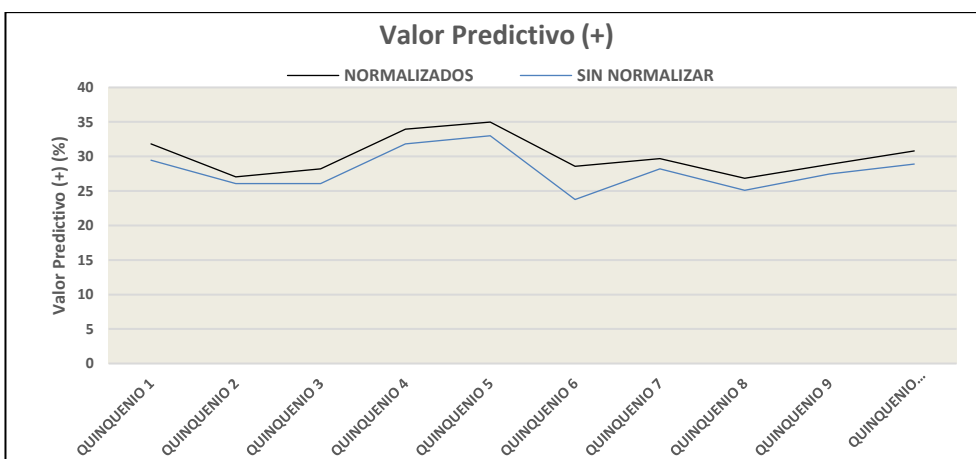


Figura 55. Valor Predictivo Positivo para Datos normalizados vs Datos sin normalizar.

Además, como ocurría en los demás casos, la línea que representa la evolución de los valores del parámetro a lo largo de los distintos quinquenios estudiados se suaviza, dando más tendencia a un valor constante que en el caso de los datos originales, lo cual presenta un aumento de calidad debido al

tratamiento de normalización.

Se trata de una evidencia clara en cuanto a la mejora que supone en el modelo de ajuste y predicción de incidencias el realizar un tratamiento de normalización estadística a los factores cuantitativos que definen los registros.

El valor máximo de dicho parámetro se alcanza en el quinquenio 5, nuevamente, con un resultado de 34.982% frente a un 33% en el caso de los datos originales.

Supone un incremento de casi un 2% en el valor predictivo con tan solo realizar un tratamiento de normalización a algunos de los factores que definen los registros.

○ Índice de Calidad

El índice de calidad se definía como la semisuma entre la sensibilidad y el valor predictivo positivo (figura 56).

La comparación entre los valores obtenidos en base a este parámetro, en función de los dos tipos de datos estudiados (normalizados y originales) presenta una evolución análoga, debido a que el incremento del parámetro sensibilidad en el caso de los datos originales se compensa con el aumento del parámetro valor predictivo positivo para el caso de los datos sometidos a tratamiento de normalización estadística.

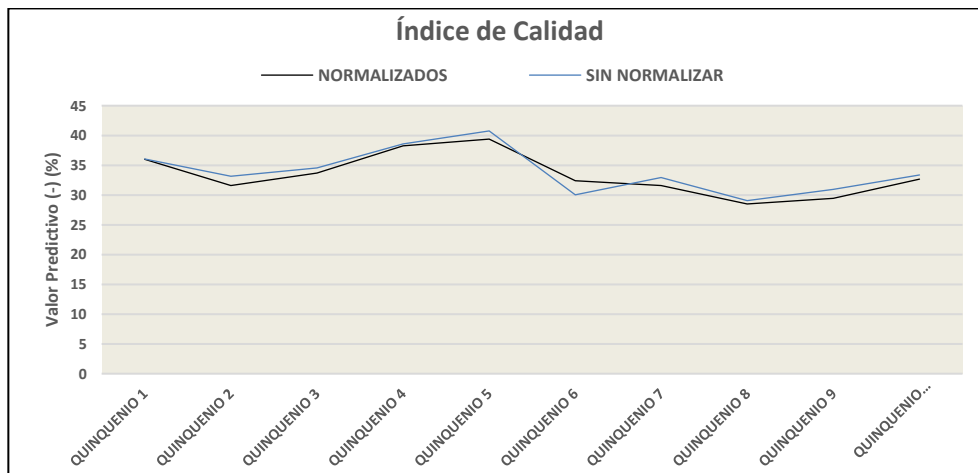


Figura 56. Índice de Calidad para Datos normalizados vs Datos sin normalizar.

3.1.5 Análisis de Coeficientes de Ajuste. Estudio de Pesos.

Como se ha comentado en el subapartado 3.1.1, la ventaja principal que proporciona el tratamiento de normalización de los factores cuantitativos es la posibilidad de comparar directamente los coeficientes de ajuste obtenidos del modelo de ajuste debido a la adimensionalización que se produce al normalizar.

A esto se dedica este apartado.

Para ello, se realizarán estudios para cada quinquenio, y posteriormente, se estudiará el caso global mediante los coeficientes obtenidos del modelo de ajuste para el periodo 2003-2017.

Se muestran a continuación el valor de los coeficientes de ajuste normalizados por quinquenios y en el período general (2003-2017).

3.1.5.1 Valores de Coeficientes de Ajuste

Se muestran en las tablas 34-37 los valores de los coeficientes de ajuste normalizados.

		Término Independiente	Diámetro	Número de acometidas	Fluctuación Presión	Presión media	Nº de Incidencias Previas Acumuladas
QUINQUENIO 2003-2008. ESTIMACIÓN 2009	Coef.con IPA	118.2064849	-0.000255848	-0.003699672	0.011510433	-0.001249763	0.2369638
	Coef.sin IPA	6.407701612	-5.28768E-05	0.000113444	-0.001057751	-6.33559E-05	0
QUINQUENIO 2004-2009. ESTIMACIÓN 2010	Coef.con IPA	159.2550239	-0.00094836	-0.007102533	0.019542954	0.002656575	0.218857795
	Coef.sin IPA	9.24148411	-4.86762E-05	4.69947E-06	-0.002112533	2.35267E-05	0
QUINQUENIO 2005-2010. ESTIMACIÓN 2011	Coef.con IPA	164.7732157	5.50009E-05	-0.007231146	0.083683395	-0.007443851	0.161038879
	Coef.sin IPA	11.11347411	-5.34412E-05	-0.000136407	-0.002442363	6.17895E-05	0
QUINQUENIO 2006-2011. ESTIMACIÓN 2012	Coef.con IPA	194.3947372	-0.000214684	-0.008507105	0.044412343	-0.002861969	0.164168024
	Coef.sin IPA	8.26619446	-5.39277E-05	-2.6149E-06	-0.000844969	-1.58927E-05	0
QUINQUENIO 2007-2012. ESTIMACIÓN 2013	Coef.con IPA	136.9869442	3.86166E-05	-0.00651887	0.039461652	-0.003698679	0.194922014
	Coef.sin IPA	8.783312483	-5.07837E-05	-5.86446E-05	-0.000394261	-8.70341E-05	0
QUINQUENIO 2008-2013. ESTIMACIÓN 2014	Coef.con IPA	100.359901	-1.74588E-05	-0.006145298	0.02867857	-0.000969453	0.196600809
	Coef.sin IPA	10.37862814	-4.62687E-05	-4.82785E-05	0.000641391	-0.000206006	0
QUINQUENIO 2009-2014. ESTIMACIÓN 2015	Coef.con IPA	122.2683526	-0.000381359	-0.005381935	0.025770639	0.000142076	0.18813061
	Coef.sin IPA	8.579213234	-4.87435E-05	-1.54969E-05	0.001294234	-0.000187684	0
QUINQUENIO 2010-2015. ESTIMACIÓN 2016	Coef.con IPA	177.5273537	-0.00068986	-0.006543725	-0.003557586	0.003224697	0.16178612
	Coef.sin IPA	6.173303555	-5.03624E-05	5.09529E-05	0.000963158	-0.000110734	0
QUINQUENIO 2011-2016. ESTIMACIÓN 2017	Coef.con IPA	196.2100585	-0.000362955	-0.004137498	-0.011797688	0.001326951	0.158490763
	Coef.sin IPA	6.013560906	-6.97186E-05	-3.89299E-05	0.000673775	-4.30506E-06	0
QUINQUENIO 2012-2017. ESTIMACIÓN 2018	Coef.con IPA	123.8372136	-0.000240393	-0.00058683	-0.003564557	-0.000964055	0.160293905
	Coef.sin IPA	6.679745257	-8.00455E-05	-8.52076E-05	0.000619773	-2.03329E-05	0
PERIODO 2003-2017. ESTIMACIÓN 2018	Coef.con IPA	49.44568868	-0.000219485	-0.001985627	0.00643715	-0.000267889	0.078067684
	Coef.sin IPA	2.809215985	-3.9845E-05	0.000136768	0.000348959	-9.28714E-05	0

Tabla 34. Coeficientes de Ajuste para Factores Cuantitativos Normalizados.

COEFICIENTES DE AJUSTE. FACTOR CUALITATIVO MATERIAL								
		FC	FD	HACH	HF	PE	PVC	VA
QUINQUENIO 2003-2008. ESTIMACIÓN 2009	Coef.con IPA	-1.897819547	-2.049581424	-1.565284201	-1.92736408	0	0	0
	Coef.sin IPA	0.034072216	0.010807042	0	0.025725302	0.007023239	0.009982026	0.033561456
QUINQUENIO 2004-2009. ESTIMACIÓN 2010	Coef.con IPA	0.120223486	-0.043512447	0.170944317	0.094186089	1.057811181	0	-0.103868469
	Coef.sin IPA	0.024182847	-0.001162463	0	0.017261031	-0.002669323	-0.000490644	0.054596131
QUINQUENIO 2005-2010. ESTIMACIÓN 2011	Coef.con IPA	-0.172471956	-0.337885975	-0.257920593	-0.207808402	0.209740611	-0.323402951	0
	Coef.sin IPA	-0.025987029	-0.057757593	-0.051818446	-0.033139961	-0.058419737	-0.057493357	0
QUINQUENIO 2006-2011. ESTIMACIÓN 2012	Coef.con IPA	0.262578841	0.105928393	0	0.240144511	0.438618549	-0.00513052	0.381397728
	Coef.sin IPA	-0.031940708	-0.063918385	-0.043870068	-0.039871877	-0.063477888	-0.061448717	0
QUINQUENIO 2007-2012. ESTIMACIÓN 2013	Coef.con IPA	0.296033202	0.149804018	0	0.275484676	0.187967784	-0.013982124	0.268299516
	Coef.sin IPA	0.011434594	-0.021583753	0	0.005845441	-0.020433048	-0.019262552	0.04583458
QUINQUENIO 2008-2013. ESTIMACIÓN 2014	Coef.con IPA	0.297520207	0.210986555	0	0.253101571	-0.018479636	0.004198824	0.201472665
	Coef.sin IPA	-0.074726962	-0.107816781	-0.084351273	-0.079764724	-0.106132802	-0.105353806	0
QUINQUENIO 2009-2014. ESTIMACIÓN 2015	Coef.con IPA	0.210824504	0.099271599	0	0.151234207	0.05745625	0	0.061031749
	Coef.sin IPA	-0.058816949	-0.090589456	-0.093070004	-0.06359432	-0.089357538	-0.099165864	0
QUINQUENIO 2010-2015. ESTIMACIÓN 2016	Coef.con IPA	0.14651326	0.096793353	0	0.091129725	0	0	0.183551128
	Coef.sin IPA	0.022392969	-0.008740686	0	0.01684242	-0.00730416	-0.017403849	0.03523158
QUINQUENIO 2011-2016. ESTIMACIÓN 2017	Coef.con IPA	0.365066168	0.336445418	0	0.324700208	0.273135815	0	0.49001774
	Coef.sin IPA	-0.013899075	-0.044374243	-0.037262812	-0.018982206	-0.043112403	-0.053558489	0
QUINQUENIO 2012-2017. ESTIMACIÓN 2018	Coef.con IPA	0.482471236	0.457461156	0	0.433659442	0.408752803	0	0.740045295
	Coef.sin IPA	-0.008350922	-0.043290445	-0.036962275	-0.018056341	-0.043174635	-0.053942936	0
PERIODO 2003-2017. ESTIMACIÓN 2018	Coef.con IPA	0.167830464	0.07314771	0.152912756	0.152245576	0.147142696	0	0.263898473
	Coef.sin IPA	0.023496539	0.002000727	0	0.017058065	0.002297194	-0.001496573	0.039630107

Tabla 35. Coeficientes de Ajuste para valores del factor cualitativo Material.

COEFICIENTES DE AJUSTE. FACTOR CUALITATIVO UBICACIÓN

		En acera	En calzada	En terrizo
QUINQUENIO 2003-2008. ESTIMACIÓN 2009	Coef.con IPA	-0.024901841	0	0.007576061
	Coef.sin IPA	0.011671061	0	0.013046301
QUINQUENIO 2004-2009. ESTIMACIÓN 2010	Coef.con IPA	0.023417634	0	0.000884347
	Coef.sin IPA	0.012077318	0	0.017496674
QUINQUENIO 2005-2010. ESTIMACIÓN 2011	Coef.con IPA	0.017507326	0	-0.05068017
	Coef.sin IPA	-0.002094801	-0.014882757	0
QUINQUENIO 2006-2011. ESTIMACIÓN 2012	Coef.con IPA	0.065114215	0.061232793	0
	Coef.sin IPA	0.012998839	0	0.01267376
QUINQUENIO 2007-2012. ESTIMACIÓN 2013	Coef.con IPA	-0.025310817	0	-0.060392688
	Coef.sin IPA	0.004766643	-0.009226353	0
QUINQUENIO 2008-2013. ESTIMACIÓN 2014	Coef.con IPA	-0.022049169	-0.053475376	0
	Coef.sin IPA	0.012878988	0	0.009335268
QUINQUENIO 2009-2014. ESTIMACIÓN 2015	Coef.con IPA	0.031886689	0	0.083489919
	Coef.sin IPA	0.011877694	0	0.007367546
QUINQUENIO 2010-2015. ESTIMACIÓN 2016	Coef.con IPA	0.069444666	0	0.212165905
	Coef.sin IPA	0.011161844	0	0.00138783
QUINQUENIO 2011-2016. ESTIMACIÓN 2017	Coef.con IPA	0.055533966	0	0.075013646
	Coef.sin IPA	0.01461702	0.000798722	0
QUINQUENIO 2012-2017. ESTIMACIÓN 2018	Coef.con IPA	0.10571937	0	0.205318711
	Coef.sin IPA	0.019817582	0.006623028	0
PERIODO 2003-2017. ESTIMACION 2018	Coef.con IPA	0.020023514	-0.017669036	0
	Coef.sin IPA	0.006392877	-0.001912331	0

Tabla 36. Coeficientes de Ajuste para valores del factor cualitativo Ubicación.

COEFICIENTES DE AJUSTE. FACTOR CUALITATIVO DISTRITO

		SE_BELLAVISTA-LA PALMERA	SE_CASCO ANTIGUO	SE_CERRO-AMATE	SE_ESTE	SE_LOCAL	SE_LOS REMEDIOS
QUINQUENIO 2003-2008. ESTIMACIÓN 2009	Coef.con IPA	0.163333458	0.163484762	0.072900805	0.232731693	-0.01700592	0
	Coef.sin IPA	0.019726313	0.021205612	0.013637683	0.035567609	0	0.00843342
QUINQUENIO 2004-2009. ESTIMACIÓN 2010	Coef.con IPA	0.152288412	0.163818665	0.024705688	0.255765591	0.051924255	0
	Coef.sin IPA	0.016595448	0.014461411	0.007389017	0.025615726	-0.00599522	0
QUINQUENIO 2005-2010. ESTIMACIÓN 2011	Coef.con IPA	0.120689694	0.110428327	0.07007344	0.26747328	0	-0.03057703
	Coef.sin IPA	0.016507392	0.012511847	0.00339607	0.024856047	-0.00628925	0
QUINQUENIO 2006-2011. ESTIMACIÓN 2012	Coef.con IPA	0.182698913	0.089861019	-0.02559028	0.169310587	0	-0.02512955
	Coef.sin IPA	0.011973498	0.012323469	0.004705505	0.023859425	-0.00808456	0
QUINQUENIO 2007-2012. ESTIMACIÓN 2013	Coef.con IPA	0.212448616	0.085952058	-0.06545025	0.151344423	-0.00189547	0
	Coef.sin IPA	0.013678186	0.013139995	0.00661348	0.02425177	-0.01035837	0
QUINQUENIO 2008-2013. ESTIMACIÓN 2014	Coef.con IPA	0.198477576	0.03416515	-0.04435048	0.106877222	0	-0.03199519
	Coef.sin IPA	0.023652286	0.023056925	0.014591021	0.036835189	0	0.010420703
QUINQUENIO 2009-2014. ESTIMACIÓN 2015	Coef.con IPA	0.198123982	0.061112863	0.022240386	0.143204772	0	0.071236938
	Coef.sin IPA	0.019193399	0.01803225	0.009138935	0.02690535	0	0.007206734
QUINQUENIO 2010-2015. ESTIMACIÓN 2016	Coef.con IPA	0.163493327	-0.01471117	-0.04281341	0.140983288	0	-0.00555781
	Coef.sin IPA	0.017782661	0.019050777	0.010649724	0.023901953	0	0.010056551
QUINQUENIO 2011-2016. ESTIMACIÓN 2017	Coef.con IPA	0.076281059	-0.06132539	-0.00869763	0.135046883	0.035423484	0
	Coef.sin IPA	0.011721278	0.010793368	0.002337246	0.012181327	-0.00940805	0
QUINQUENIO 2012-2017. ESTIMACIÓN 2018	Coef.con IPA	0.062133686	0.042936538	-0.01308531	0.122545296	0	0.020717556
	Coef.sin IPA	0.022115668	0.016750824	0.011738788	0.02256519	0	0.009345477
PERIODO 2003-2017. ESTIMACION 2018	Coef.con IPA	0.102299072	0.048132675	-0.004149	0.065754362	0	0.006317327
	Coef.sin IPA	0.011210073	0.011173692	0.008621358	0.01691677	0	0.005819756

		SE_MACARENA	SE_NERNIVON	SE_NORTE	SE_SAN PABLO-SANTA JUSTA	SE_SUR	SE_TRIANA
QUINQUENIO 2003-2008. ESTIMACIÓN 2009	Coef.con IPA	-0.03594107	0.039612672	0.00119296	0.041438793	0.004794689	0.019024319
	Coef.sin IPA	0.013222861	0.014855555	0.01405389	0.015415122	0.018219618	0.017549112
QUINQUENIO 2004-2009. ESTIMACIÓN 2010	Coef.con IPA	-0.03973303	0.071380119	0.011376655	0.086157803	0.008627825	0.046267818
	Coef.sin IPA	0.003415627	0.008426135	0.006615514	0.006831067	0.010502084	0.007961428
QUINQUENIO 2005-2010. ESTIMACIÓN 2011	Coef.con IPA	-0.03987409	0.070647668	-0.00982605	0.07853593	0.020528758	-0.00457127
	Coef.sin IPA	-0.00125394	0.006225776	0.003276212	0.00220129	0.007009357	0.009369346
QUINQUENIO 2006-2011. ESTIMACIÓN 2012	Coef.con IPA	-0.06645827	0.033107775	-0.05696269	0.009901343	-0.00815496	0.030037409
	Coef.sin IPA	-0.00340692	0.006123208	0.001161402	0.003354456	0.005463212	0.005607784
QUINQUENIO 2007-2012. ESTIMACIÓN 2013	Coef.con IPA	-0.06358494	0.110354465	0.061138171	0.060913753	0.039819792	0.026726645
	Coef.sin IPA	-0.00229884	0.005708393	-5.766E-05	0.003062111	0.005162262	0.007081725
QUINQUENIO 2008-2013. ESTIMACIÓN 2014	Coef.con IPA	-0.03749593	0.123495622	0.08854121	0.065990897	0.084661535	0.063864684
	Coef.sin IPA	0.007046506	0.016132443	0.009105418	0.01386864	0.015537605	0.016823302
QUINQUENIO 2009-2014. ESTIMACIÓN 2015	Coef.con IPA	0.022668467	0.110861183	0.0910574	0.020749412	0.093242415	0.152904805
	Coef.sin IPA	0.002955379	0.013217514	0.006503736	0.010830578	0.012739998	0.014105415
QUINQUENIO 2010-2015. ESTIMACIÓN 2016	Coef.con IPA	-0.06973391	0.008937014	0.071996471	0.007187147	-0.01219957	0.031688462
	Coef.sin IPA	0.004868885	0.014292539	0.007318692	0.01074121	0.013316014	0.014877291
QUINQUENIO 2011-2016. ESTIMACIÓN 2017	Coef.con IPA	-0.06672708	-0.02653105	0.063565581	-0.0414776	-0.01199183	-0.02052156
	Coef.sin IPA	-0.00402981	0.007653157	0.000174675	0.0038252	0.003307149	0.005758654
QUINQUENIO 2012-2017. ESTIMACIÓN 2018	Coef.con IPA	-0.04021432	0.022789905	-0.01352423	0.016237596	0.026639994	-0.01384173
	Coef.sin IPA	0.003157839	0.016058548	0.007969533	0.011213039	0.010203715	0.015485512
PERIODO 2003-2017. ESTIMACION 2018	Coef.con IPA	-0.04877798	0.04546229	-0.00524743	0.010532535	0.002772082	0.009252229
	Coef.sin IPA	0.005603816	0.008635661	0.00553859	0.007597341	0.009138642	0.010089233

Tabla 37. Coeficientes de Ajuste para valores del factor cualitativo Distrito.

3.1.5.2 Representación de la evolución de Coeficientes de Ajuste normalizados por quinquenios.

En este subapartado se representa gráficamente los valores de los coeficientes de ajuste mostrados en las tablas 34-37 (figuras 57-65).

Para ello, primeramente, se ha representado la evolución, por un lado, de los factores cuantitativos en conjunto (Diámetro, Número de Acometidas, Presión Media y Fluctuación de Presión).

A su vez, se mostrarán los valores de los coeficientes obtenidos, por un lado, teniendo en cuenta las IPA (Incidencias Previas Acumuladas), y sin tenerlas en cuenta, para comprobar cómo cambian los valores.

Posteriormente se representarán los valores de los coeficientes asignados a los distintos atributos de los factores cualitativos, estudiando uno a uno cada uno de los factores.

Por último, se muestra la comparación de evolución de todos los factores en conjunto en el llamado periodo general (periodo que abarca todos los años de los que se disponen datos, desde 2003 hasta 2017 como fase de ajuste, y 2018 como año de predicción).

Para poder relacionarlos se ha aplicado un procedimiento para calcular el efecto que tiene variar un factor en concreto sobre el parámetro PF.

El procedimiento se diferencia dependiendo de si se trata un factor cuantitativo o cualitativo:

- En el caso de tratarse de un factor cuantitativo, una vez hallado el valor del coeficiente de ajuste, este se multiplica por la diferencia existente entre los valores máximo y mínimo del factor tras el procedimiento de normalización.
- Si es un factor cualitativo, el efecto se obtiene restando los coeficientes máximo y mínimo que la regresión ha asignado a cada una de las categorías del factor. De este modo, se garantiza una correcta relación entre coeficientes.

- **Factores normalizables.**

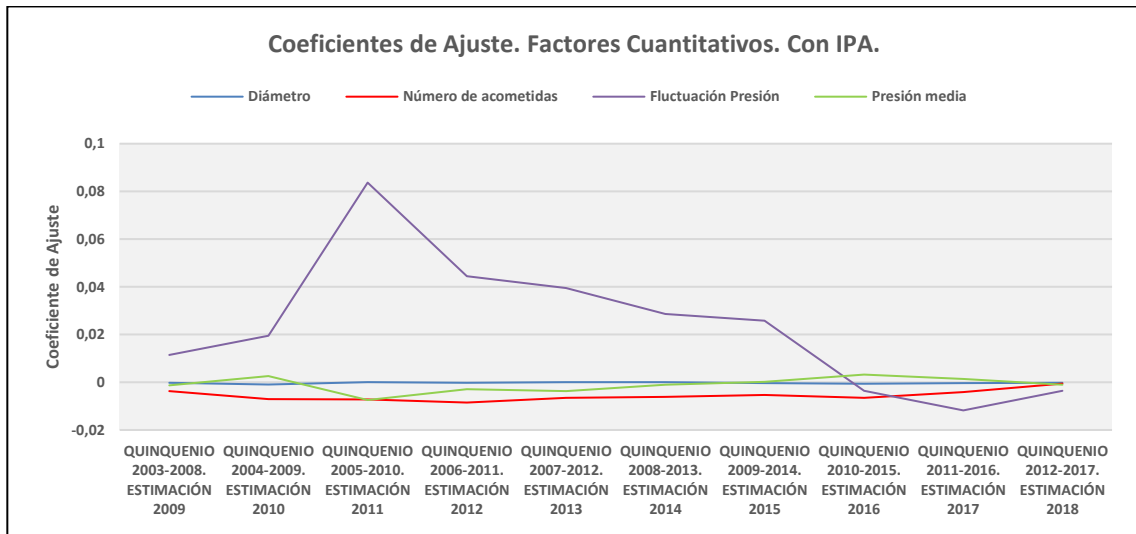


Figura 57. Evolución de coeficientes de Ajuste Normalizados para Factores Cuantitativos Normalizados. Con IPA.

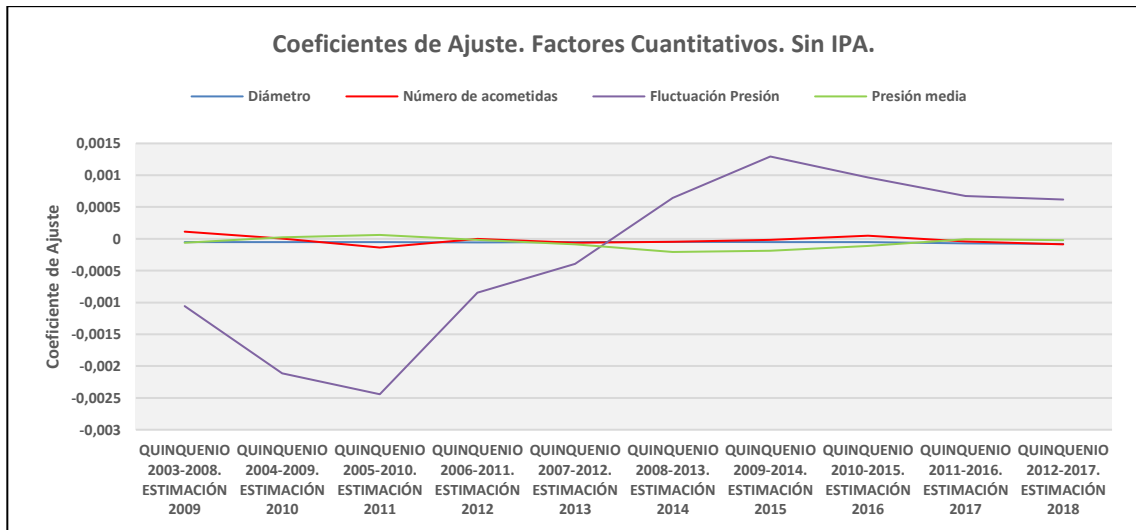


Figura 58. Evolución de coeficientes de Ajuste Normalizados para Factores Cuantitativos Normalizados. Sin IPA.

- *Regresión para las familias con IPA > 0*

Sobre la evolución de los coeficientes que multiplican a los factores cuantitativos teniendo en cuenta las IPA se observa lo siguiente:

- En el caso del factor Diámetro siempre toma valores negativos, o valores positivos, pero muy cercanos a la media, lo cual indica que sería un factor candidato a salir del análisis de regresión.
- En el caso del factor Fluctuación de Presión, se trata de un factor que cuyo coeficiente de ajuste está generalmente por encima de la media, por lo tanto, es un factor que aumenta el riesgo de aparición de incidencias, menos en los últimos quinquenios, donde decae hasta alcanzar valores negativos.
- El coeficiente del factor Presión Media, presenta valores positivos y negativos con la misma frecuencia, manteniéndose el valor de este coeficiente muy cercano a cero en todo momento. Se puede concluir por lo tanto que no tiene un efecto muy considerable sobre el modelo de predicción.
- En cuanto al factor Número de Acometidas, este en todo momento se encuentra por debajo de la media, es decir, toma valores negativos. Este efecto se manifiesta en la predicción de forma que, al tomar valores negativos y positivos (efecto de la normalización), serían más propensos según la regresión a sufrir incidencias aquellas familias que tuvieran menos acometidas.

- *Regresión para las familias con IPA=0*

Sobre la evolución de los valores de los coeficientes sin tener en cuenta el IPA, se extraen las siguientes observaciones:

- En cuanto al factor Diámetro, toma valores negativos en la mayoría de quinquenios muy cercanos a cero.
 - En el caso del coeficiente asignado al factor Fluctuación de Presión, y sin considerar las IPA, toma valores negativos en los primeros quinquenios, y a partir del quinquenio 6 comienza a tomar valores positivos. Esto se traduce en que, en los primeros quinquenios, las familias con una fluctuación de presión pequeña registran más incidencias que las que presentan más fluctuación de presión, mientras que, en la segunda mitad de los quinquenios estudiados, el contexto cambia, y las familias que están sometidas a mayores cambios de presión tienen más incidencias registradas en los quinquenios correspondientes a la fase de ajuste para esos años.
 - En referencia al factor Presión Media, de forma parecida a lo que ocurre con el Diámetro, sus valores fluctúan muy cercanos al valor medio entre positivos y negativos. Sería, por lo tanto, otro factor candidato a salir del ajuste.
 - En cuanto al factor Número de Acometidas, en este caso no siempre está por debajo de la media (valor 0), lo cual indica que aquellas familias con menor número de acometidas están sometidas según el modelo de regresión a un riesgo mayor de aparición de incidencias que las que tienen un elevado número de acometidas.
- **Factores no normalizables**

❖ Factor Número de Incidencias Previas acumuladas

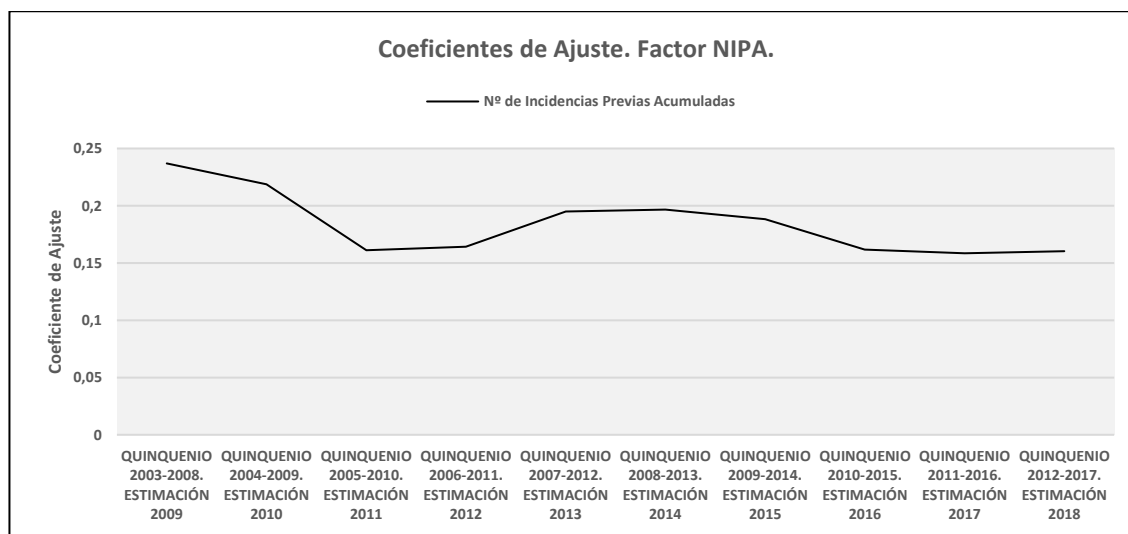


Figura 59. Evolución de Coeficientes de Ajuste No Normalizados para Factor Cuantitativo Número de Incidencias Previas.

En el caso del factor Número de Incidencias Previas Acumuladas (figura 59), es importante destacar que este factor no se ha podido normalizar manualmente debido a que todos los tratamientos que se le aplican a este factor se realizan de forma automática mediante procesos internos del software de ajuste y predicción. Los valores del coeficiente en su evolución a lo largo de los quinquenios representan los valores más altos de entre todos los coeficientes que acompañan a los factores cuantitativos. Este hecho indica que se trata en este caso del factor que mejor se comporta frente a la predicción de incidencias, pues, cuanto inferior sea a la media (dícese por media el valor 0), más se opone a la predicción de incidencias.

❖ Factor Material

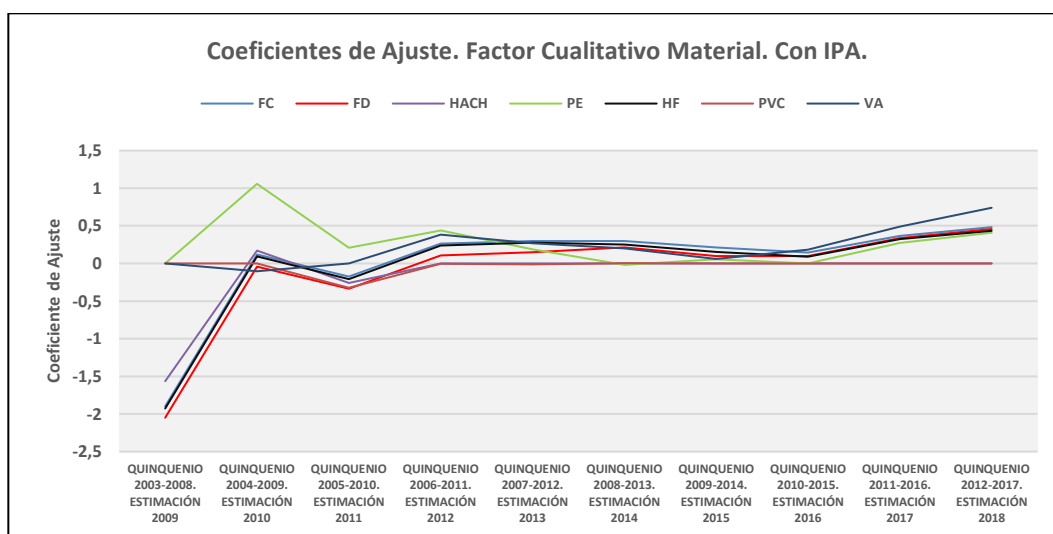


Figura 60. Evolución de Coeficientes de Ajuste No Normalizados para Factor Cualitativo Material. Con IPA.

- Regresión para las familias con $IPA > 0$ (figura 59)

Las observaciones realizadas sobre la evolución de los coeficientes de ajuste para el factor Material, teniendo en cuenta las IPA, indican que la mayoría de materiales tienen comportamientos similares, como el HACH, el FD, el FC y el HF, presentando diferencias entre ellos, pero poco significativas.

En los primeros quinquenios, manteniéndose los demás factores constantes, el cambio de un tipo de material a otro produce el aumento o disminución de riesgo de sufrir una incidencia según el modelo de regresión. Así, por ejemplo, si una familia que en principio está conformada por FD en el segundo quinquenio, pasase a estar conformada por PE, los coeficientes de ajuste muestran que dicha familia pasaría de tener un riesgo bajo de sufrir incidencias en términos del factor material (FD, con coeficiente negativo), a tener un riesgo más elevado de sufrirlas (PE con coeficiente positivo).

Esta explicación se puede extrapolar a todos los coeficientes.

El material que más se aleja de esta evolución es el material PE. En los primeros quinquenios este material presenta coeficientes de ajuste positivos, lo cual indica que, en los primeros quinquenios estudiados, una familia conformada por el material PE tiene más riesgo de sufrir incidencia que si estuviera conformada por otro material, manteniendo los demás factores constantes.

- Regresión para las familias con $IPA = 0$ (figura 60)

En este caso, la mayoría de materiales en general presentan coeficientes negativos bastante alejados de la media, es decir, se oponen a la aparición de incidencias. Al igual que en el caso anterior presentan evoluciones paralelas con ciertas variaciones entre los valores de los coeficientes de ajuste.

En este caso, la diferencia entre los coeficientes de ajuste es más notable. El material que más influencia tiene en la aparición de incidencias es el material VA, frente a los demás.

Las familias que menos probabilidad tienen de sufrir una incidencia son las conformadas por PVC y PE en el caso estudiado, ya que presentan los valores de coeficientes más bajos.

Según los resultados, si una familia está conformada por PVC, y se produjera un hipotético cambio de material, a FC, manteniendo los demás factores sin modificaciones, en el quinquenio 5, por ejemplo, la familia que, en términos del factor material, tenía un riesgo muy bajo de sufrir incidencia (dado el signo negativo del coeficiente de ajuste asignado al material PVC en el quinquenio en cuestión), pasaría a tener un riesgo más elevado debido a que el coeficiente de ajuste asignado al material FC es positivo en dicho quinquenio.

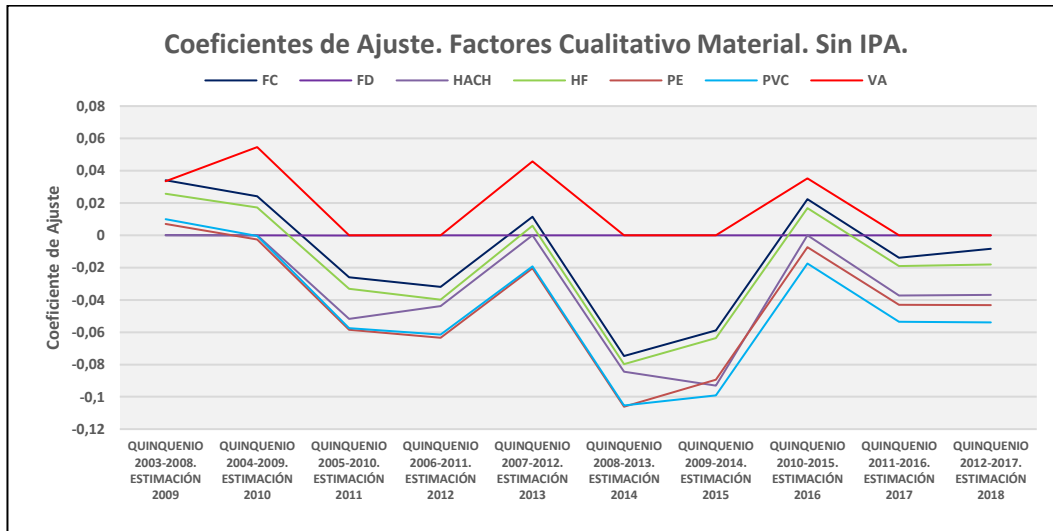


Figura 61. Evolución de Coeficientes de Ajuste No Normalizados para Factor Cualitativo Material. Sin IPA.

❖ Factor Ubicación

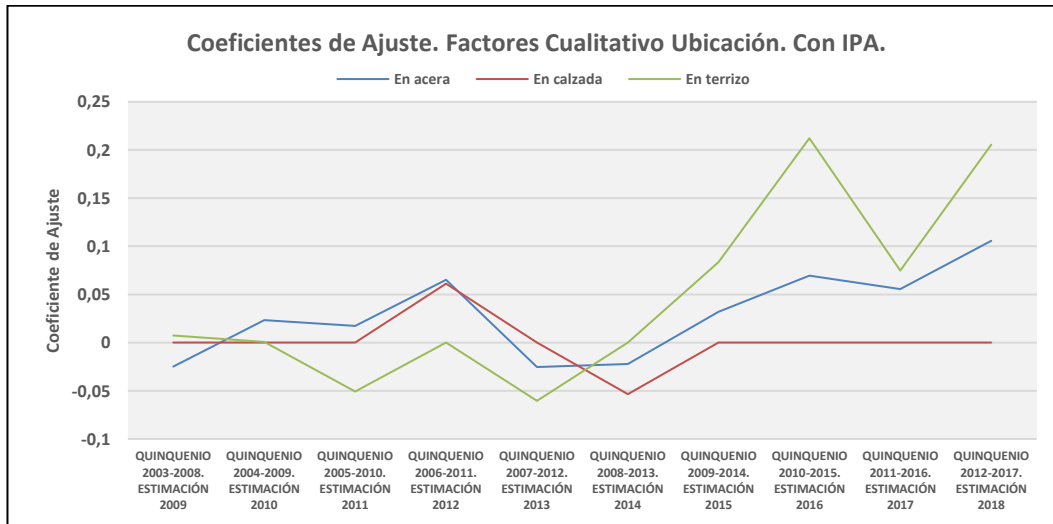


Figura 62. Evolución de coeficientes de Ajuste Normalizados para Factor Cualitativo Ubicación. Con IPA.

- *Regresión para las familias con $IPA > 0$ (figura 62)*

En los primeros quinquenios, las familias que más riesgo de sufrir incidencia presentan son las localizadas En Acera. A medida que los quinquenios avanzan las familias con localización En Terrizo presentan más riesgo de incidencia que las ubicadas En Acera.

Las familias localizadas En Calzada en todo momento tienen, aun siendo el valor de los coeficientes de ajuste positivos, menos riesgo de sufrir incidencia que las demás.

- *Regresión para las familias con $IPA = 0$ (figura 63)*

En el caso del factor Ubicación, según los coeficientes de ajuste obtenidos mediante el modelo de regresión, las familias que estén localizadas en la ubicación En Acera tienen un riesgo más elevado de sufrir una incidencia que las familias localizadas En Calzada.

Las familias que estén localizadas En Terrizo presentan en la mayoría de quinquenios riesgo de sufrir incidencias, pero no es tan alto (según el factor Ubicación) que en el caso de las localizadas En Acera.

En este caso, por ejemplo, una familia que se localice en la ubicación En Calzada en el quinquenio que

abarca desde 2008 hasta 2013, no contabilizaría incidencias por parte de este factor en el modelo de regresión debido a que el valor del coeficiente de ajuste es 0, mientras que si la familia estuviese localizada En Acera, registraría en torno a 0.012 incidencias debido al valor de su coeficiente de ajuste.

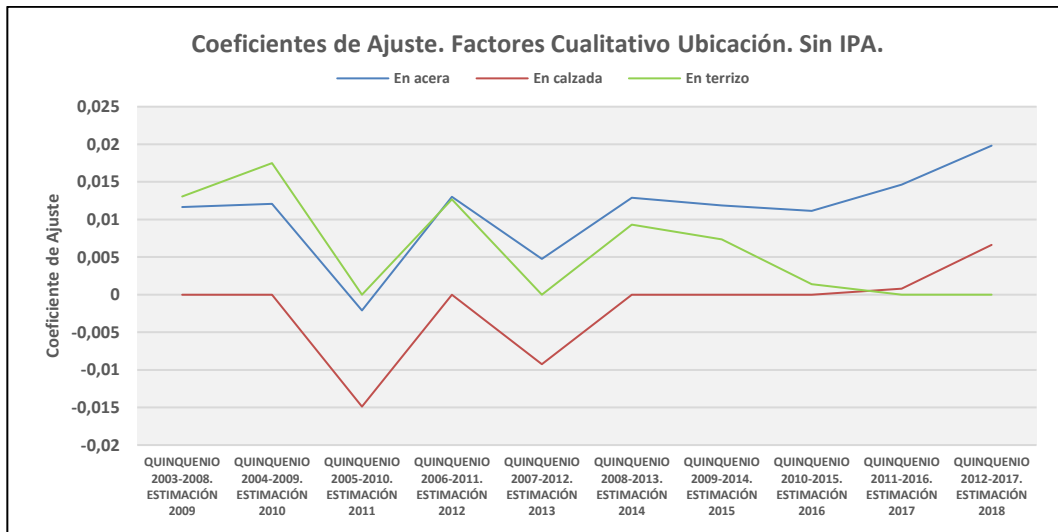


Figura 63. Evolución de coeficientes de Ajuste Normalizados para Factor Cualitativo Ubicación. Sin IPA.

❖ Factor Distrito

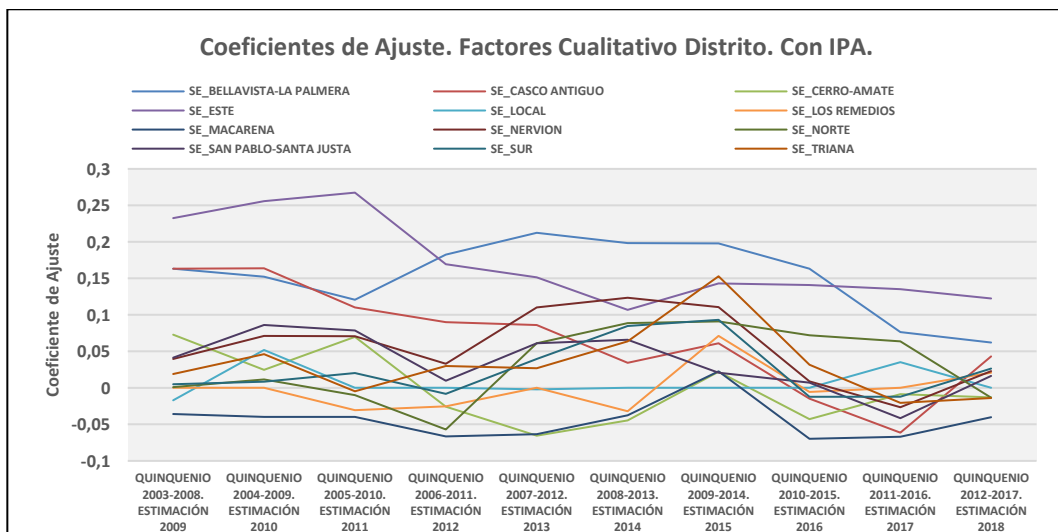


Figura 64. Evolución de coeficientes de Ajuste Normalizados para Factor Cualitativo Distrito. Con IPA.

- Regresión para las familias con $IPA > 0$

En este caso, la evolución de los valores que toman los distintos coeficientes es un poco caótica.

En los primeros quinquenios, las familias localizadas en distritos como SE_CASCO ANTIGUO o SE_ESTRE presentan un riesgo mucho más elevado que las demás de sufrir incidencias (en términos del factor distrito). A medida que avanzan los quinquenios, las familias localizadas en el distrito SE_LOCAL comienzan a presentar más riesgo de incidencia, y las anteriormente comentadas disminuyen.

Por lo general, la mayoría de coeficientes presentan valores positivos, por lo que se puede decir que generalmente el Distrito es un factor que aumenta el riesgo de incidencia a las familias.

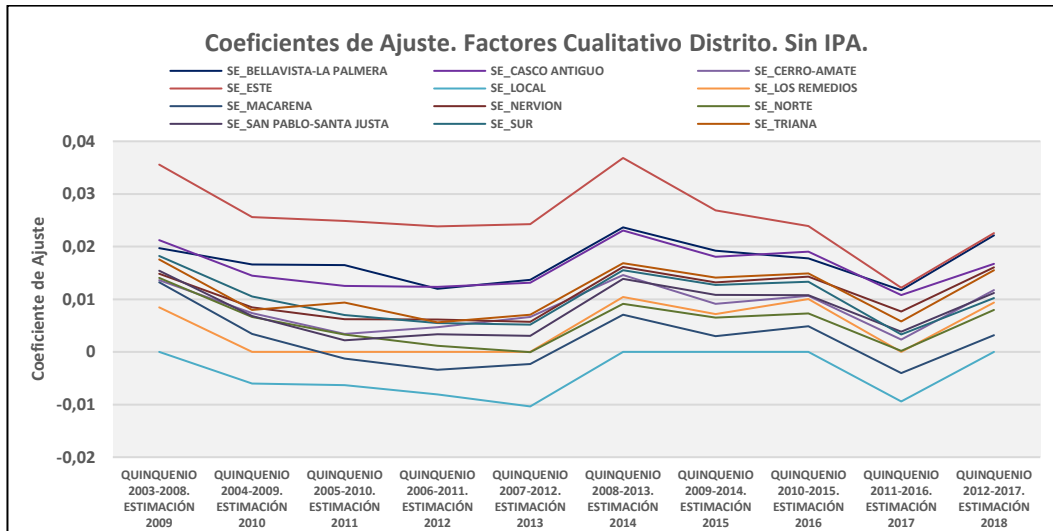


Figura 65. Evolución de coeficientes de Ajuste Normalizados para Factor Cualitativo Distrito. Sin IPA.

- Regresión para las familias con IPA=0

En este caso, los coeficientes presentan evoluciones similares a lo largo de los quinquenios, pero los valores que toman son distintos en cada uno. Las familias que menos riesgo presentan de sufrir incidencias son las localizadas en el distrito SE_LOCAL (puede ser debido al bajo número de familias que se registran con este distrito). Por lo general, las familias localizadas en el distrito SE_ESTE tienen más riesgo de incidencia que las localizadas en los demás distritos.

La mayoría de distritos presentan coeficientes de ajuste con valores similares, que giran en torno al 0.01 y 0.02, manteniendo valores positivos.

En este caso, igual que con la consideración de IPA, el distrito es un factor que atribuye riesgo de incidencia a las familias, dados los resultados de los coeficientes de ajuste obtenidos.

• Relación general entre Coeficientes de Ajuste. Cuantitativos y Cualitativos.

En la tabla 38 se muestran los valores obtenidos tras aplicación del procedimiento descrito al principio de este subapartado, basado en el efecto que tendría sobre el parámetro PF una modificación de los factores cuantitativos y cualitativos, con el fin de poder comparar sus pesos entre sí. Estos resultados aparecen representados gráficamente en las figuras 66 y 67, respectivamente.

	Diámetro	Número de acometidas	Fluctuación Presión	Presión media	FC	FD	HACH	HF
Con IPA	-0.002977999	-0.035433311	0.039459728	-0.006085767	0.044290203	0.019303569	0.040353443	0.040177375
Sin IPA	-0.000540621	0.002440616	0.002139121	-0.002109807	0.000966335	8.22833E-05	0	0.000701542

PE	PVC	VA	En acera	En calzada	En terrizo	SE_BELLAVISTA-LA PALMERA	SE_CASCO ANTIGUO	SE_CERRO-AMATE
0.038830733	0	0.069642404	0.000754737	-0.000665991	0	0.015455042	0.007271743	-0.000626818
9.4476E-05	-6.15491E-05	0.001629855	5.30942E-05	-1.58823E-05	0	0.000189638	0.000189023	0.000145846

SE_ESTE	SE_LOCAL	SE_LOS REMEDIOS	SE_MACARENA	SE_NERVION	SE_NORTE	SE_SAN PABLO-SANTA JUSTA	SE_SUR	SE_TRIANA
0.009933975	0	0.000954403	-0.007369233	0.006868309	-0.000792766	0.001591224	0.000418798	0.001397799
0.000286177	0	9.84515E-05	9.47985E-05	0.000146087	9.36951E-05	0.000128522	0.000154596	0.000170677

Tabla 38. Coeficientes para comparación de Factores Cuantitativos y Cualitativos normalizados.

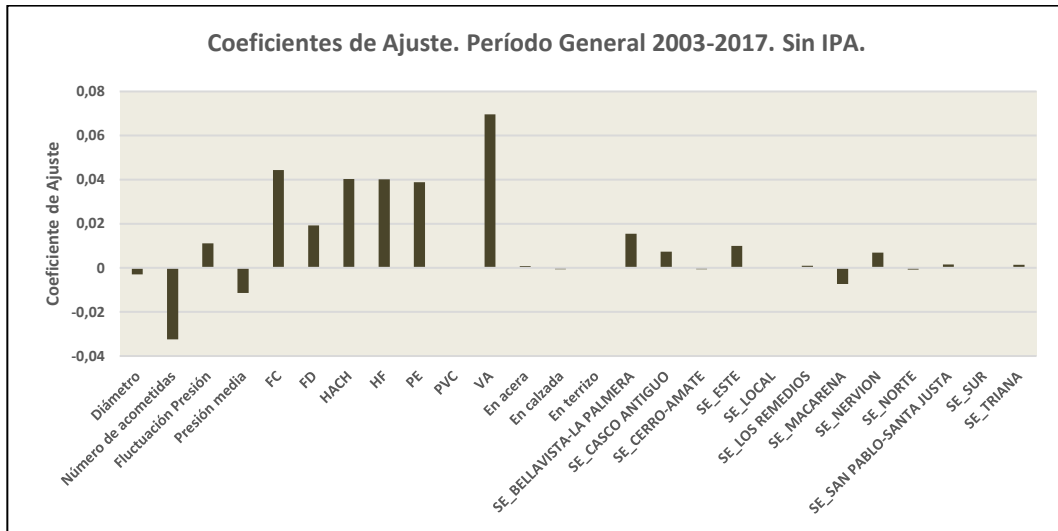


Figura 66. Relación entre coeficientes de ajuste a nivel General. Con IPA.

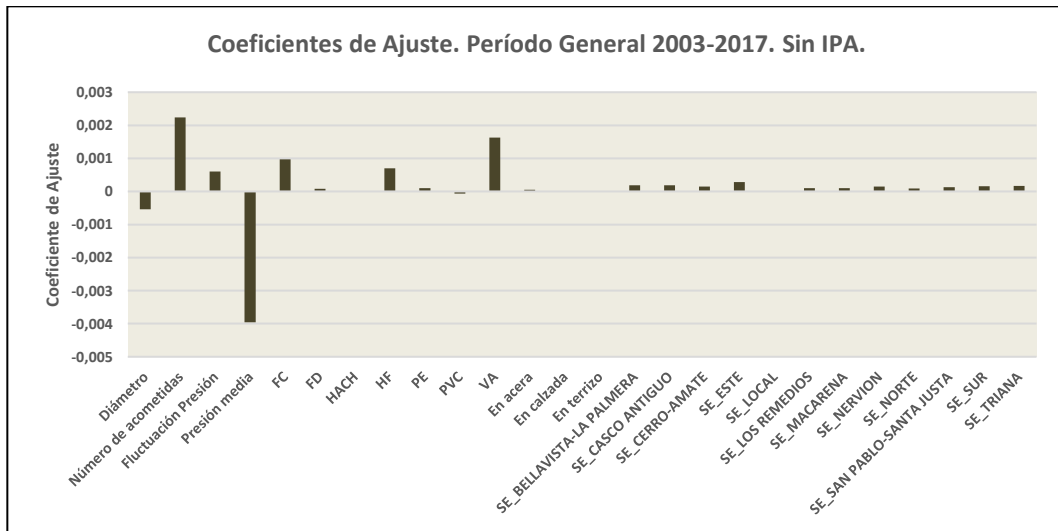


Figura 67. Relación entre coeficientes de ajuste a nivel General. Con IPA.

- *Regresión para las familias con IPA > 0*

Frente a la relación conjunta de los factores tanto cuantitativos como cualitativos, se observa que en general, los cualitativos tienen más peso sobre el ajuste que los cuantitativos. En particular el factor material, que para todos los valores que puede tomar presenta coeficientes de ajuste positivos bastante altos, atribuyendo así a las familias riesgo de sufrir incidencias.

En el caso de los factores cuantitativos, tienen más peso en el ajuste los factores Número de acometidas, que reduce el riesgo de incidencia siempre y cuando el número de acometidas normalizado esté por encima de 0, y la presión media.

La fluctuación de presión mantiene un coeficiente positivo, es decir, cuanto mayor sea la fluctuación de presión asociada a una familia, mayor será el riesgo de sufrir incidencia.

Estas conclusiones se pueden extrapolar a todos los factores que no se han comentado.

- *Regresión para las familias con IPA = 0*

En este caso los factores que más peso tienen en el ajuste son el número de acometidas, la fluctuación de presión y el material FC. Aquellas familias que presenten un alto número de acometidas, una presión media normalizada por debajo de 0, y estén constituidas por FC, presentaran el riesgo más alto de incidencia.

3.2 Constatación de relación entre PF y Probabilidad de Fallo

El penúltimo de los objetivos planteados en este TFG, el cual se busca en este apartado es corroborar la relación entre el coeficiente PF calculado con la regresión y la probabilidad de fallo real de la tubería.

3.2.1 Descripción del Desarrollo

Para el desarrollo de los resultados, se han incluido todos los factores de los que se tiene disponibilidad en el modelo.

Los filtros que se aplican sobre los registros para el desarrollo de este apartado son:

- Selección de tramos que cuentan con una longitud igual o superior a 3 m.
- Selección de tramos que pertenecen a la RSA
- Selección de tramos que pertenecen al municipio de Sevilla.

En cuanto a los demás factores, se incluyen todos los registros independientemente del valor que adopten estos factores (Material, Diámetro, Año de Instalación, etc.).

El funcionamiento del proceso llevado a cabo para corroborar la relación consiste en:

1. Mediante el modelo de regresión lineal multiparámetro implementado en el “Software de Ajuste y Predicción” se obtienen las puntuaciones asignadas del coeficiente PF a las familias sometidas a predicción en el periodo en cuestión.
2. Se establecen una serie de intervalos en función del coeficiente PF. Los intervalos que se han establecido aumentan con paso de 0.1 del coeficiente PF.
3. Posteriormente, se calculan las incidencias reales que han sido registradas en las familias en función del intervalo al que pertenecen en base a la puntuación asignada del coeficiente PF.
4. Para calcular la Probabilidad de Fallo real de las familias, se realiza la división entre el número de incidencias reales acumuladas en un intervalo delimitado por el PF, y el total de incidencias reales ocurridas en el período.
5. Por último, se representan mediante diagrama de barras la relación existente entre la Probabilidad de Fallo calculada y los distintos intervalos delimitados del coeficiente PF.

Se comprobará el modelo para cuatro períodos concretos, elegidos arbitrariamente. Los periodos para comprobación que han sido elegidos han sido los siguientes:

- **Período 1:** la fase de ajuste abarca desde el año 2003 hasta el año 2010, realizando la predicción para 2011.
- **Período 2:** la fase de ajuste abarca desde los años 2008 hasta el año 2015, realizando la predicción para 2016.
- **Período 3:** la fase de ajuste abarca desde el año 2010 hasta el año 2017, realizando la predicción para el año 2018 (último año del que se tiene registro de incidencias).
- **Período 4:** supone el periodo de ajuste más amplio. La fase de ajuste abarca desde el año 2003 hasta el año 2017, realizando la predicción para el año 2018.

Un ejemplo de la estructura de la tabla de resultados obtenidos para cada período de ajuste y predicción se muestra en la tabla 39. Se obtiene una hoja de registros existentes en el año o periodo de comprobación, exactamente igual que los utilizados como datos para ejecutar el modelo, con la adición de tres columnas nuevas:

- **IR (Incidencias Reales):** este parámetro representa el número de incidencias reales registradas en el registro correspondiente al tramo o familia sobre el que se ha realizado la predicción.

- **PF (Previsión de Fallo):** parámetro que asigna un nivel de riesgo a un cierto tramo o familia. Resultado del cálculo del número de incidencias estimadas.
- **RIESGO asociado al PF (Previsión de Fallo):** este parámetro muestra las distintas clasificaciones (MEDIO, ALTO, MUY ALTO) de los tramos o familias en función del factor PF asignado.

Con el PF ya calculado, y teniendo la clasificación, se comparará la existencia o no de posibles incidencias registradas en el periodo o año en cuestión para la predicción frente al valor de Previsión de Fallo.

Para la comparativa se presentan en un gráfico de barras, el porcentaje de familias que han sufrido una o más incidencias, en función de su clasificación por el factor PF, frente al total de familias con incidencias registradas en el período o año de ajuste.

Familia	Material	Diámetro	Año Instalación	Indicador de Borrado	IR	PF	RIESGO asociado al PF
1	HF	200	1950	NO	0	0.696	MEDIO
3	HF	200	1950	NO	0	0.509	MEDIO
4	HF	100	1950	NO	0	0.727	MEDIO
5	HF	100	1950	NO	1	0.749	MEDIO
6	HF	80	1950	NO	0	0.73	MEDIO
7	HF	80	1950	NO	0	0.749	MEDIO
8	HF	150	1950	NO	0	0.507	MEDIO
9	HF	125	1950	NO	0	0.507	MEDIO
10	HF	125	1950	NO	0	0.715	MEDIO
11	HF	200	1980	NO	0	0.509	MEDIO
13	HF	200	1980	NO	0	0.726	MEDIO
14	HF	100	1980	NO	0	0.767	MUY ALTO
15	HF	80	1980	NO	0	0.505	MEDIO
17	HF	150	1980	NO	0	0.733	MEDIO
19	HF	150	1980	NO	0	0.503	MEDIO
20	HF	200	1985	NO	0	0.504	MEDIO
22	HF	200	1985	NO	0	0.715	MEDIO
23	HF	100	1985	NO	0	0.772	MUY ALTO
24	HF	100	1985	NO	0	0.749	MEDIO

Tabla 39. Estructura de Resultados para el factor PF.

3.2.2 Resultados

- **Primer Período. Fase Ajuste: 2003-2010. Verificación 2011.**

En la tabla 40 aparecen los valores de incidencias reales, y de la Probabilidad de Fallo para el periodo de ajuste que abarca desde el año 2003 hasta el año 2010, siendo el año para la verificación, el 2011, en función de intervalos para el factor PF (Previsión de fallo) que aparecen en la columna de la izquierda. Estos intervalos varían cada 0.1 unidades de PF.

PERIODO 2003-2010. PREDICCIÓN 2011		
PROBABILIDAD DE FALLO FRENTE A PF		
PF	INCIDENCIAS REALES	PROBABILIDAD
PF<=0.1	0	0.000
PF<=0.2	0	0.000
PF<=0.3	0	0.000
PF<=0.4	0	0.000
PF<=0.5	3	0.012
PF<=0.6	54	0.219
PF<=0.7	74	0.300
PF<=0.8	160	0.648
PF<=0.9	212	0.858
PF<=1	247	1.000

Tabla 40. Probabilidad de Fallo frente a Previsión de Fallo. Período de Ajuste 2003-2010. Predicción 2011.

Como se observa en los valores de la columna situada en la derecha de la tabla, la probabilidad de fallo se mantiene nula hasta que se alcanza un intervalo de $PF \leq 0.5$, donde, a partir de dicho intervalo, la probabilidad de fallo aumenta a medida que aumenta el intervalo de PF, acercándose a la unidad.

Las variaciones con respecto a los demás intervalos se mantienen, si no constantes, proporcionales, no incrementándose de un intervalo a otro en más de 0.35 puntos.

En la figura 68, se muestra la evolución gráfica de los valores de probabilidad frente a PF ilustrados en la tabla 40.

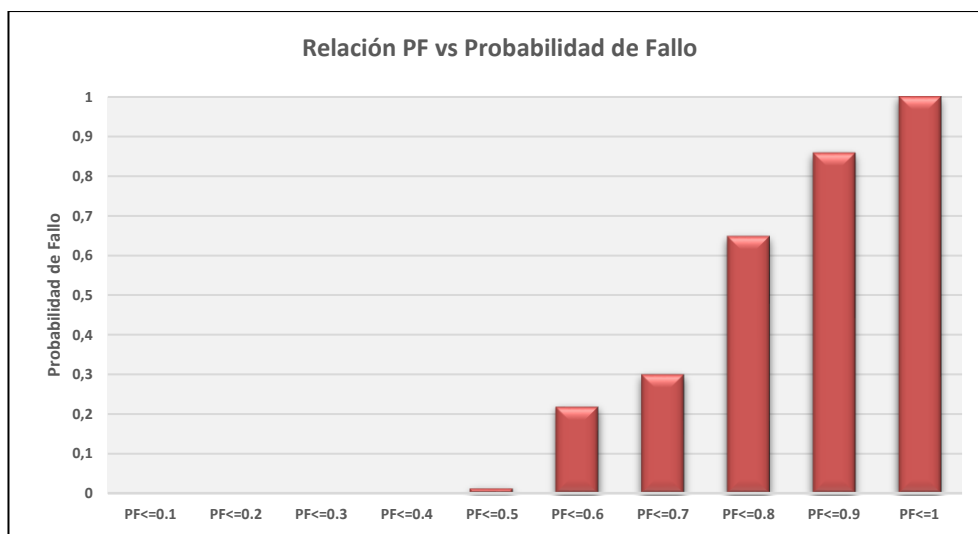


Figura 68. Relación factor PF frente a Probabilidad de Fallo. Período de Ajuste 2003-2010. Predicción 2011.

Se observa pues, que no ha fallado ninguna familia con una puntuación del PF inferior a 0.4.

Los valores de probabilidad de fallo van aumentando a medida que aumenta el intervalo del factor PF, lo cual demuestra la correcta correlación del índice PF con la probabilidad de fallo

Para el caso de las familias que cuentan con un valor de PF menor o igual que 0.6, la probabilidad de fallo es de 0.219, aumentando en 0.207 puntos respecto al intervalo anterior.

El incremento más notable para este periodo se observa en el paso del intervalo $PF \leq 0.7$ al intervalo $PF \leq 0.8$, siendo el incremento de 0.348.

Las familias con un valor de PF asignado inferior o igual a 0.8 tienen asignada una probabilidad de fallo ya considerable, de valor 0.648.

▪ **Segundo Período. Fase Ajuste: 2008-2015. Verificación 2017.**

En la tabla 41 se observan los valores de incidencias reales para el periodo de ajuste 2008-2015, y año de predicción 2017, junto con la probabilidad de fallo de cada rango de familias incluidas en los intervalos indicados para el factor PF.

PERÍODO 2008-2015. PREDICCIÓN 2016		
PROBABILIDAD DE FALLO FRENTE A PF		
PF	INCIDENCIAS REALES	PROBABILIDAD
PF ≤ 0.1	0	0
PF ≤ 0.2	0	0
PF ≤ 0.3	0	0
PF ≤ 0.4	0	0
PF ≤ 0.5	8	0.035874439
PF ≤ 0.6	60	0.269058296
PF ≤ 0.7	72	0.322869955
PF ≤ 0.8	160	0.717488789
PF ≤ 0.9	202	0.905829596
PF ≤ 1	223	1

Tabla 41. Probabilidad de Fallo frente a Previsión de Fallo. Período de Ajuste 2008-2015. Predicción 2017.

Los valores de la tabla 41, se representan gráficamente en la figura 69, de forma análoga al periodo de estudio anterior.

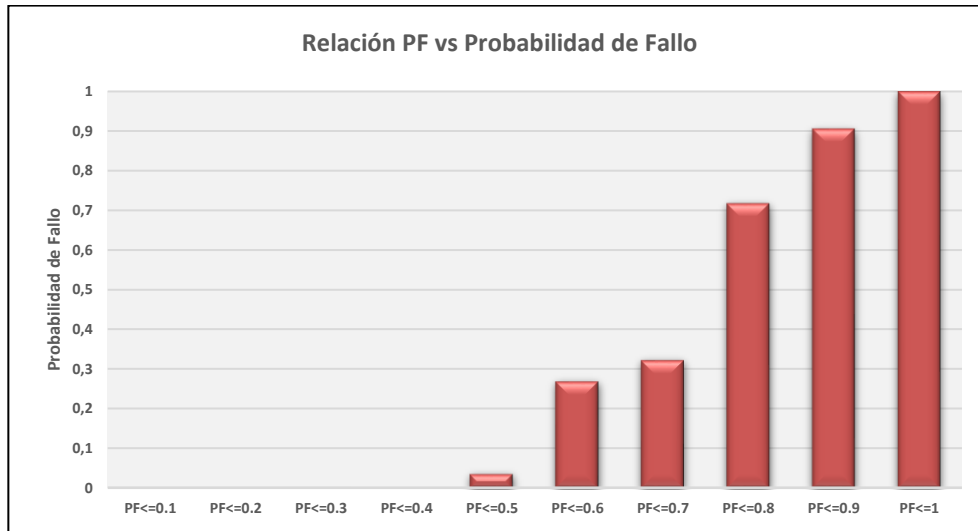


Figura 69. Relación factor PF frente a Probabilidad de Fallo. Periodo de Ajuste 2008-2015. Predicción 2017.

En el periodo de ajuste 2008-2015, se puede observar una evolución de la Probabilidad de Fallo semejante al periodo anterior.

En este caso, tampoco se ha dado ninguna incidencia en familias con una puntuación de PF inferior a 0.4.

La probabilidad de fallo comienza a tomar valores distintos de cero, en las familias que tienen asignado un valor del factor PF igual a 0.5, y a partir de dicho intervalo comienza a tomar valores positivos hasta que se alcanza el valor unidad para aquellas familias con un valor del factor PF entre 0.9 y 1.

En este periodo, aumentan las probabilidades de fallo correspondientes a los distintos intervalos del factor PF, en todos los casos.

Para el intervalo $PF \leq 0.5$, se produce un aumento de 0.0238, pasando de un valor de 0.012 que se tenía en el periodo anterior, a un valor de 0.358 actual.

En el caso del periodo $PF \leq 0.6$, igualmente, se produce un incremento de 0.05 en la probabilidad de fallo.

Para el intervalo que abarca a aquellas familias con un valor del factor PF inferior o igual a 0.7, el valor de la probabilidad de fallo alcanza el 0.3228, 0.0228 superior al caso anterior.

El incremento más notable se produce en el intervalo $PF \leq 0.8$, donde se pasa de un valor en el periodo anterior de 0.648, a un valor actual de 0.717. Este representa un valor muy importante de la probabilidad de fallo.

Para el intervalo $PF \leq 0.9$, se produce un caso muy curioso, y es que el valor del factor PF (0.9), es exactamente igual a la probabilidad de fallo asignada a las familias pertenecientes a ese tramo, de valor también 0.9.

Obviamente, al igual que ocurre en el caso anterior, aquellas familias que tengan asignadas un valor de PF igual que 1, cuentan con una probabilidad de fallo máxima, de valor unidad.

▪ **Tercer Período. Fase Ajuste: 2010-2017. Verificación 2018.**

En la tabla 42 aparecen los valores de probabilidad de fallo y número de incidencias reales en función de los intervalos estudiados para el factor Previsión de Fallo, asignados al periodo de estudio que incluye en el modelo de ajuste y regresión los datos más recientes.

El periodo de ajuste abarca desde el año 2010, hasta el año 2017, siendo el año de predicción y verificación el año 2018.

Posteriormente, aparece representada la evolución de la probabilidad de fallo en la figura 70, respecto a los intervalos del factor PF.

PERIODO 2010-2017. PREDICCIÓN 2018		
PROBABILIDAD DE FALLO FRENTE A PF		
PF	INCIDENCIAS REALES	PROBABILIDAD
PF<=0.1	0	0
PF<=0.2	0	0
PF<=0.3	0	0
PF<=0.4	0	0
PF<=0.5	2	0.008403361
PF<=0.6	59	0.24789916
PF<=0.7	69	0.289915966
PF<=0.8	178	0.74789916
PF<=0.9	224	0.941176471
PF<=1	238	1

Tabla 42. Probabilidad de Fallo frente a Previsión de Fallo. Período de Ajuste 2010-2017. Predicción 2018.

Al igual que ocurría en los dos periodos anteriores, no existe ninguna familia con una puntuación de PF inferior a 0.4 con incidencias registradas.

En el caso de los intervalos $PF \leq 0.6$ y $PF \leq 0.7$, el valor de la probabilidad de fallo asignado a esas familias es muy parecido, de valores 0.2478 y 0.2899, respectivamente.

Se produce un fuerte incremento del valor de la probabilidad de fallo para aquellas familias que tienen asignado el valor del factor PF, en el rango o intervalo $PF \leq 0.8$, donde pasa a tomar la probabilidad de fallo, un valor de 0.7478, produciéndose un incremento respecto al intervalo anterior de 0.4589. Incremento muy considerable.

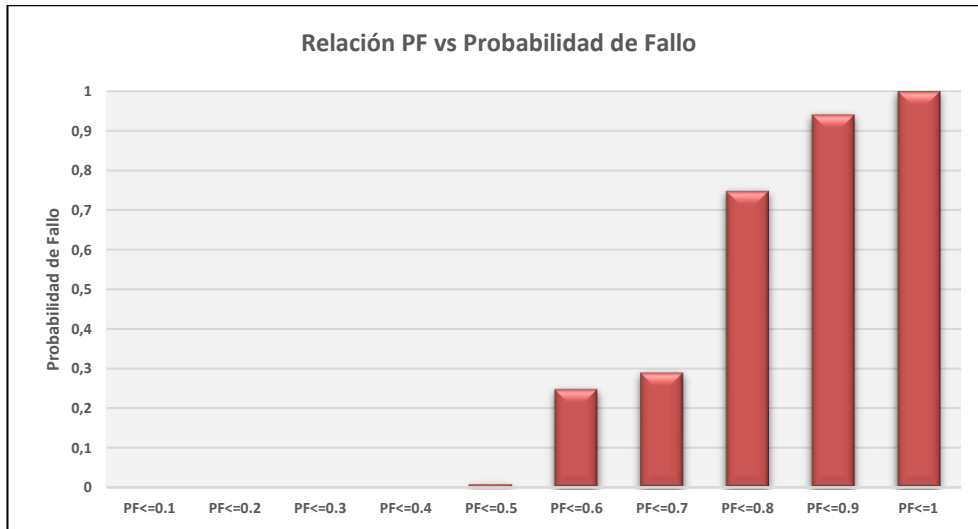


Figura 70. Relación factor PF frente a Probabilidad de Fallo. Período de Ajuste 2010-2017. Predicción 2018.

▪ **Período General. Fase de Ajuste: 2003-2017. Predicción 2018.**

En la tabla 43 se muestran los resultados de probabilidad e incidencias reales frente a distintos valores del factor PF asignado a las familias.

La evolución de la probabilidad de fallo frente al factor PF aparece posteriormente representada gráficamente en la figura 71.

El periodo de ajuste en este caso es el más amplio de los cuatro estudiados. Abarca desde el año 2003 (primer año en el que comienzan los registros extraídos de las bases de datos de EMASESA) hasta el año 2017. El año de predicción y verificación es 2018.

PERIODO 2003-2017. PREDICCIÓN 2018

PROBABILIDAD DE FALLO FRENTE A PF

PF	INCIDENCIAS REALES	PROBABILIDAD
PF<=0.1	0	0
PF<=0.2	0	0
PF<=0.3	0	0
PF<=0.4	0	0
PF<=0.5	4	0.016806723
PF<=0.6	46	0.193277311
PF<=0.7	59	0.24789916
PF<=0.8	180	0.756302521
PF<=0.9	220	0.924369748
PF<=1	238	1

Tabla 43. Probabilidad de Fallo frente a Previsión de Fallo. Período de Ajuste 2003-2017. Predicción 2018.

En este caso, se aprecia una evolución de la probabilidad de fallo un poco más irregular que en los casos anteriores.

Se distinguen 3 rangos claramente diferenciados:

- Rango Nulo: engloba a todas las familias que tienen asignado un valor del factor PF hasta el intervalo $PF \leq 0.4$, y que no han sufrido incidencias.
- Rango de Probabilidad de Fallo Baja: engloba a aquellas familias que tienen asignada una probabilidad de fallo entre $PF \leq 0.5$, y $PF \leq 0.7$, donde la evolución sufre incrementos graduales, siendo el valor de la probabilidad de fallo en $PF \leq 0.5$, de valor 0.0168, y siendo máximo en el rango en cuestión para el intervalo $PF \leq 0.7$, de valor 0.2478.
- Rango de Probabilidad de Fallo Alta: este rango engloba desde las familias que tienen asignadas valores del factor PF entre aquellos mayores que 0.7 y 1. En este caso aparecen probabilidades bastante altas, siendo para $PF \leq 0.8$, de valor 0.756, y para $PF \leq 0.9$, de valor 0.924.

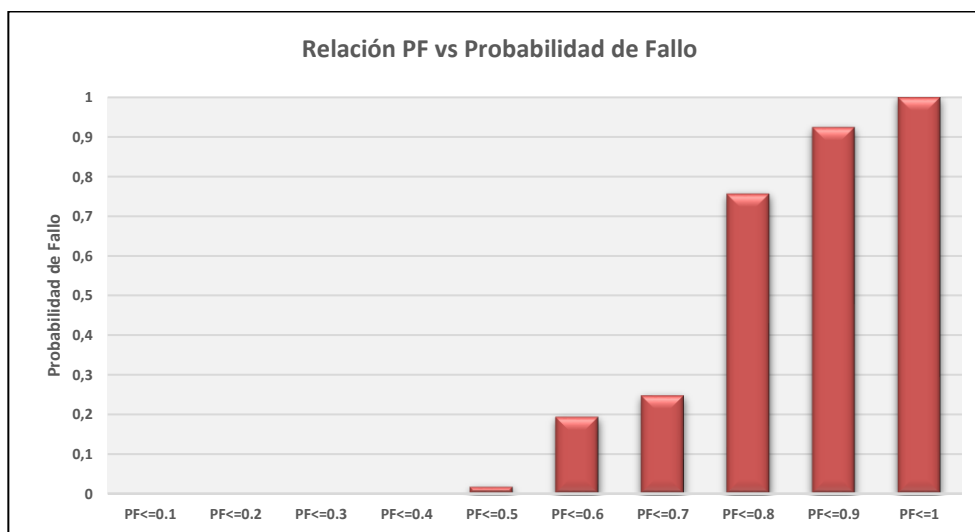


Figura 71. Relación factor PF frente a Probabilidad de Fallo. Período de Ajuste 2003-2017. Predicción 2018.

4 CONCLUSIONES Y DESARROLLOS FUTUROS

En este capítulo se describen en términos generales las conclusiones obtenidas tras el desarrollo del TFG, y se proponen algunas actuaciones futuras derivadas de la interpretación de dichas conclusiones.

El capítulo se estructura en 3 apartados:

- 1) El primer apartado recoge las conclusiones obtenidas a partir de la observación de resultados sobre el estudio de factores. Este a su vez se estructura en 2 subapartados. El primer subapartado describe las conclusiones obtenidas sobre el estudio de correlación de factores desarrollado a partir del uso de las Tablas de Contingencia, y el segundo subapartado describe las conclusiones derivadas de la observación de los resultados obtenidos tras el proceso de filtrado de *outliers*.
- 2) El segundo apartado resume las conclusiones obtenidas tras la observación de los resultados obtenidos en el capítulo 3. Este a su vez, también se compone de 2 subapartados. Un primer subapartado en el que se describen las conclusiones derivadas del tratamiento de normalización a los datos de entrada, y el segundo subapartado describe las conclusiones obtenidas tras el estudio de correlación del parámetro PF, frente a la Probabilidad de Fallo.
- 3) Por último, el tercer apartado se dedica a describir algunas propuestas de desarrollos futuros que se han obtenido a partir de las conclusiones.

4.1 Conclusiones del Capítulo 2. Análisis de Factores

4.1.1 Conclusiones sobre la Correlación de Factores Cualitativos

- **Relación Material vs Ubicación.** El resultado del análisis de correlación sobre estos dos factores ha dado una relación mediante el coeficiente K. Pearson de 0.11. Se trata de una relación bastante baja, por lo que, aunque no se cumpla la hipótesis de independencia de χ^2 , se concluye con que estos dos factores están débilmente relacionados.
- **Relación Material vs Distrito.** En este caso, la relación establecida por el coeficiente K. Pearson es de 0.3618. Se trata de una relación mayor que en el caso Material-Ubicación, por lo que se va a concluir con que estos factores están poco relacionados.
- **Relación Distrito vs Ubicación:** la relación que establece el coeficiente K. Pearson entre estos dos materiales es de 0.1791. Aunque según el método de χ^2 no se cumple la no relación entre estos dos factores, el resultado mediante el coeficiente K. Pearson establece que están ligeramente relacionados.

4.1.2 Conclusiones sobre filtrado de *Outliers*.

Tras examinar los resultados obtenidos al realizar los correspondientes análisis de calidad frente al intento de eliminación de *outliers*, se han obtenido las siguientes justificaciones:

- Tras la comparación de los resultados obtenidos para el parámetro Sensibilidad, se concluye con que el filtrado mejora la sensibilidad, que aumenta de 40.5% a 42.78%.
- El parámetro Valor Predictivo Positivo cambia respecto a la Sensibilidad, existiendo una mejora de los resultados para los registros modificados en los primeros quinquenios estudiados, y posteriormente, presentando mejor resultados el uso de registros originales.
- Por último, el parámetro Índice de Calidad (el cual se corresponde con la semisuma de la Sensibilidad y el Valor Predictivo Positivo), no presenta grandes variaciones entre ambos casos. En los primeros quinquenios mejoran los resultados obtenidos en base a los registros modificados, y en quinquenios posteriores mejoran los resultados para los datos originales, pero en general, la

variación de resultados no es muy importante.

El coste que ha supuesto el filtrado, en volumen de datos eliminados ha sido muy elevado (del orden del 40%), mientras que en las técnicas de ML (automáticas) se suelen eliminar entre un 5% y un 10%, lo que hace cuestionable el interés del filtrado tal y como se ha propuesto.

4.2 Conclusiones del Capítulo 3. Tratamiento de Datos y Contraste de Resultados.

4.2.1 Conclusiones Acerca del Tratamiento de Datos.

4.2.1.1 Conclusiones Procedimiento de Normalización

Tras aplicar el tratamiento de normalización a los factores cuantitativos y comparar los resultados para los factores sin normalizar frente a los factores normalizados, se aprecia que el único parámetro que decae con el uso de los factores cuantitativos es el parámetro sensibilidad. Aun así, la mayor diferencia entre ambos no supera el 4%.

Para el parámetro Valor Predictivo Positivo, el cual se trata de uno de los parámetros más importantes para el análisis de calidad de la regresión, se produce también un incremento por parte de los datos con factores normalizados frente a los datos originales, además de incrementarse la suavidad en su evolución a lo largo de los quinquenios estudiados. En el quinquenio 6 se da la mayor diferencia entre ambos, de un 4.808% de incremento del Valor Predictivo Positivo, el cual se trata de un incremento bastante importante.

Por último, el Índice de Calidad, mantiene una evolución casi paralela entre ambos grupos de datos, dándose algunos picos (máximos y mínimos) más pronunciados en el caso del conjunto de datos sin normalizar. Para este caso, la evolución a lo largo de los distintos quinquenios estudiados para el conjunto de datos normalizados presenta una mayor suavidad, lo que da sensación de mayor calidad del ajuste por parte de dicho conjunto.

En definitiva, se justifica en base a los resultados presentados, que aplicar un tratamiento de normalización a los factores cuantitativos produce una mejora de los resultados.

4.2.1.2 Conclusiones sobre Coeficientes de Ajuste.

- Factores Normalizables

- *Regresión para las familias con $IPA > 0$*

En orden de influencia, el factor que tiene el coeficiente mayor es el factor Fluctuación de Presión, seguido del factor, seguido del factor Número de Acometidas, Presión Media, y Diámetro.

Mientras que el coeficiente derivado del factor Fluctuación de Presión toma valores positivos en la mayoría de quinquenios, es decir, el número de incidencias asociadas a este factor aumenta cuanto más aumenta el valor del factor, en el caso de los demás factores no se genera este efecto, si no que ocurre el efecto contrario. A medida que disminuye el valor de los factores, aumenta el valor de incidencias asociadas a estos.

- *Regresión para las familias con $IPA = 0$*

En este caso, el coeficiente del factor Fluctuación de Presión se observa que es un poco sensible al rango de datos, por lo que, puede ser preferible estudiarlo en un rango de años más amplio para comprobar su evolución. En definitiva, requeriría más estudio.

En cuanto a los demás factores, presentan efectos similares entre ellos, con valores de sus coeficientes muy cercanos a cero.

- Factores no normalizables

- *Regresión para las familias con $IPA > 0$*

En orden de influencia, el factor que más relevancia tiene en el modelo de regresión es el factor Material,

con órdenes de magnitud entre -2 y 1, con un promedio cercano a 0.5, seguido por el factor Número de Incidencias Previas Acumuladas, con un coeficiente del orden de 0.2. Posteriormente se encuentra el factor Ubicación, y con menos influencia el factor Distrito.

- *Regresión para las familias con $IPA=0$*

El factor que más influencia presenta en el ajuste es el factor Material, que presenta valores de entre 0.05 y -0.1, seguido de los factores Ubicación y Distrito, que presentan de media, efectos similares, obteniendo coeficientes en el ajuste que varían entre -0.015 y 0.03.

4.2.2 Conclusiones sobre la Comparación Factor PF-Probabilidad de Fallo.

Los resultados obtenidos en el apartado 3.2.2 muestran que, para todos los periodos estudiados, el parámetro PF mantiene una relación con la Probabilidad de Fallo muy estrecha.

Para $PF \leq 0.5$ la probabilidad de fallo es muy baja, a partir de ese valor, y hasta $PF \leq 1$ la relación con la probabilidad de fallo puede considerarse casi-lineal.

4.3 Desarrollos Futuros

4.3.1 Correlación de Factores Cualitativos

Existen gran variedad de herramientas para el análisis de correlación tanto para factores cuantitativos como cualitativos. En este TFG sólo se ha empleado un método con los factores cualitativos, pero este análisis puede extenderse a factores cuantitativos (definiendo categorías) y también existen métodos más sofisticados como, por ejemplo: el método T de Student, el coeficiente de correlación de Pearson, o el coeficiente de regresión.

4.3.2 Filtrado de Outliers

Los criterios utilizados para detectar *outliers* en este TFG han demostrado ser excesivamente severos, implicando la eliminación de un número muy importante de registros. Por ello, sería recomendable utilizar algún procedimiento que permita controlar el número de registros eliminados, como la eliminación con alguna técnica de ML de un cierto porcentaje de los registros más alejados del hiperplano de ajuste, o la definición de criterios para cada factor menos restrictivos.

4.3.3 Procedimiento de normalización y los coeficientes de ajuste

En el Análisis de Factores Cuantitativos, se ha puesto de manifiesto la mejora de las predicciones con la normalización. Sería interesante añadir al software una macro que realice dicho procedimiento para los factores cuantitativos (incluidos aquellos que no se han podido analizar en este TFG. Ello permitiría extender la normalización a los factores: longitud, edad, e incidencias previstas acumuladas.

La aplicación a los factores cualitativos no es trivial, pero existen propuestas de métodos para cuantificación de variables cualitativas, con el fin de poder aplicar un tratamiento de normalización, como por ejemplo [6].

4.3.4 Relación del índice PF con la Probabilidad de Fallo

Aprovechando la relación quasi-lineal entre el factor PF y la Probabilidad de Fallo, sería interesante disponer de una función que las relacione explícitamente, de forma que se disponga de una medida con un significado claro del riesgo de incidencia. No obstante, el número de ensayos realizado es relativamente pequeño, y se requerirían más evidencias de esta relación. Además, este análisis de relación junto con la interpretación gráfica proporciona información suficiente para designar unos nuevos límites en la clasificación a posteriori de asignar la puntuación, para afinar, en la exactitud de las predicciones.

5 BIBLIOGRAFÍA

- [1] FIUS, «Investigación aplicada al análisis del comportamiento y durabilidad de los materiales de las redes de abastecimiento y saneamiento para la optimización de las herramientas de GPI de EMASESA.,» Universidad de Sevilla, Sevilla, 2019.
- [2] FIUS, «Herramienta de Priorización de Actuaciones sobre una Red de Distribución de aguas basada en el Índice de Fallo. Aplicación a la Red de Abastecimiento.,» Universidad de Sevilla, Sevilla, 2019.
- [3] FIUS, «Herramienta de Priorización de Actuaciones sobre una Red de Distribución de aguas basada en el Índice de Fallo. Manual de Usuario.,» Universidad de Sevilla, Sevilla, 2019.
- [4] Z. Hernández Martín, Métodos de Análisis de Datos: Apuntes, La Rioja: Servicio de Publicaciones, Universidad de la Rioja, 2012.
- [5] Y. Dodge, The Oxford Dictionary of Statistical Terms, International Statistical Institute, 2003.
- [6] M. Vásquez, G. Ramírez y A. Carmadiel, Un procedimiento para la cuantificación de variables cualitativas basado en los principios de la promediación recíproca y el análisis de homogeneidad, Caracas: Universidad Central de Venezuela, 2001.

6 ANEXO. TABLAS DE RESULTADOS

6.1 Tablas de Resultados Estudio de Factores.

DISTRITOS	Red secundaria de abastecimiento	Red arterial y de transporte de abastecimiento	TODOS
AG_ALCALA DE GUADAIRA	273.354	15.052	288.407
AG_LA GALBANA	0.017	1.592	1.608
AG_LAS ENCINAS	3.61	0	3.61
AG_LOCAL	0	0	0
AG_PINOS DEL NEVERO	6.328	0	6.328
AG_RANILLA-P.I. EL SOMBRERON	0.916	3.78	4.696
AG_RUSTICA NORTE	5.603	7.933	13.537
AG_SANTA GENOVEVA	2.869	0	2.869
AG_SEVILLA GOLF	8.356	0	8.356
AG_TORREQUINTO	4.623	0.932	5.555
AG_ZONA INDUSTRIAL	149.213	34.795	184.008
AR_ALCALA DEL RIO	57.44	23.716	81.156
AR_EL VIAR	4.361	1.498	5.859
AR_ESQUIVEL	5.594	4.806	10.4
AR_ILIPA MAGNA	2.776	0.004	2.78
AR_SAN IGNACIO DEL VIAR	5.889	1.071	6.96
BU_BURGUILLOS	0	0.855	0.855
CAR_LOCAL	0	0	0
CAR_TORREPALMA	8.959	0	8.959
CA_ZONA INDUSTRIAL	11.652	0	11.652
CA_CAMAS	57.738	3.889	61.627
CA_CARAMBOLO	10.47	3.203	13.673
CA_OESTE	1.523	0	1.523
CA_PANOLETA	16.794	5.147	21.941
CO_CORIA DEL RIO	111.791	5.85	117.641
DH_ADRIANO	1.098	6.359	7.457
DH_CONDEQUINTO	22.737	0.127	22.864
DH_DOS HERMANAS	0	0.681	0.681
DH_DOS HERMANAS CENTRO ESTE	42.038	2.853	44.891
DH_DOS HERMANAS CENTRO OESTE	69.974	3.755	73.729
DH_DOS HERMANAS ESTE	53.353	4.965	58.317
DH_DOS HERMANAS NORTE	73.798	18.135	91.933
DH_DOS HERMANAS OESTE	53.266	5.924	59.189
DH_DOS HERMANAS SUR	42.632	4.205	46.838
DH_ENTRENUCLEOS	58.749	12.737	71.486
DH_FUENTE DEL REY	18.359	4.502	22.861
DH_LA ISLA	42.733	13.146	55.878
DH_LOCAL	0	0	0
DH_MONTEQUINTO	102.636	14.348	116.984
DH_PITAMO-HIPODROMO	0.611	6.398	7.009
DH_PUNTA DEL VERDE	1.732	0	1.732
EG_EL GARROBO	8.449	2.997	11.447
EG_LOCAL	0	0	0
ER_EL RONQUILLO	19.719	5.252	24.971
ER_LOCAL	0	0	0
GE_GELVES	0.028	6.731	6.759
GU_LOCAL	0	0	0
LP_LA PUEBLA DEL RIO	48.704	1.019	49.724
MA_EL TORREON	6.618	0.099	6.717
MA_LOCAL	0	0	0
MA_MAIRENA DEL ALCOR	124.567	14.571	139.137
PA_PALOMARES_RIO	0.414	2.592	3.005
RI_EL GORDILLO	3.189	0	3.189
RI_ESPARTALES_AEROPOLIS	11.365	8.103	19.468
RI_LA RINCONADA	54.11	8.728	62.838
RI_LOCAL	0	0	0
RI_NUEVA JARILLA	11.149	0.686	11.835
RI_SAN JOSE DE LA RINCONADA	118.481	21.165	139.645
RI_TARAZONA	13.092	5.805	18.897
SA_SANTIPONCE	1.518	0.532	2.051
SE_BELLAVISTA-LA PALMERA	131.077	29.572	160.649
SE_CASCO ANTIGUO	125.923	13.736	139.659
SE_CERRO-AMATE	186.787	10.937	197.723
SE_ESTE	201.282	28.018	229.3
SE_LOCAL	0.367	0	0.367

SE_LOS REMEDIOS	53.181	10.698	63.879
SE_MACARENA	102.965	8.901	111.866
SE_NERVION	97.077	5.92	102.997
SE_NORTE	194.874	23.16	218.034
SE_SAN PABLO-SANTA JUSTA	125.597	9.86	135.457
SE_SUR	140.856	21.602	162.457
SE_TRIANA	81.451	20.058	101.509
SJ_LOCAL	0	0.073	0.073
SJ_SAN JUAN DE AZNALFARACHE	60.514	10.658	71.171
TODOS	3256.947	483.73	3740.678

Tabla 44. Longitud de Red frente a Distritos.

n°acom	LONGITUD FRENTE A Nº DE ACOMETIDAS	
	RSA	RAA
0	419470.53	424348.9
1	169470.456	31604.148
2	129924.215	17180.166
3	90052.682	13766.913
4	81338.138	5662.587
5	75454.735	2431.091
6	62663.066	200.725
7	59108.662	3012.159
8	55852.673	277.512
9	78780.28	1822.786
10	52914.193	0
11	56391.249	347.288
12	57700.409	409.106
13	51488.969	0
14	54334.437	336.517
15	37912.049	1391.636
16	40618.334	116.073
17	37531.756	0
18	42969.698	1072.048
19	37994.788	502.407
20	42751.574	242.502
21	33235.166	0
22	31213.163	253.464
23	37303.629	919.695
24	30644.903	0
25	33592.084	0
26	39049.265	0
27	35645.249	0
28	28940.723	0
29	27617.732	396.532
30	19854.094	199.93
31	29834.839	0
32	21742.173	0
33	30383.486	0
34	18813.073	0
35	24905.694	0
36	19775.074	307.251
37	14476.931	0
38	20165.643	0
39	21714.871	0
40	16121.923	0
41	19753.891	0
42	16223.524	0
43	14931.23	0
44	28881.062	0
45	22035.828	0
46	14011.945	0
47	17281.971	0
48	15570.007	0
49	20984.181	0
50	18549.893	0
51	21025.746	0
52	18201.768	0
53	15198.105	0
54	14922.695	0
55	11090.223	0
56	11116.169	0
57	10376.766	0
58	15995.275	0

126	3128.214	0
127	1943.306	0
128	8442.483	0
129	6423.188	0
130	4208.799	0
131	4665.611	0
132	1109.885	0
133	4900.744	0
134	2818.563	0
135	1916.874	0
136	2407.138	0
137	728.791	0
138	775.79	0
139	2149.659	0
140	7842.153	0
141	1451.353	0
142	2953.645	0
143	1964.558	0
145	2072.559	0
146	2715.014	0
147	4319.675	0
148	4905.47	0
149	1834.74	0
150	4230.52	0
151	667.999	0
152	4147.787	0
153	8160.023	0
154	1316.692	0
156	2045.262	0
157	3014.753	0
158	3758.507	0
159	1459.827	0
160	1779.864	0
161	4087.046	0
163	2940.288	0
164	1897.16	0
165	4050.008	0
166	4217.661	0
167	5028.547	0
168	4164.98	0
170	5774.431	0
171	3120.554	0
172	4474.367	0
173	3695.68	0
174	1753.169	0
175	3427.587	0
176	3557.992	0
177	3387.497	0
178	1211.919	0
179	827.341	0
180	6401.703	0
181	3997.502	0
182	856.778	0
184	3974.466	0
185	1449.308	0
187	1199.884	0
188	1813.618	0
189	2738.903	0
192	5456.567	0

59	11904.885	0
60	13389.268	0
61	14090.794	0
62	16764.478	0
63	11104.588	0
64	9653.297	0
65	16521.4	0
66	11480.794	0
67	9507.504	0
68	9236.768	0
69	8925.681	0
70	11938.263	0
71	14011.745	0
72	18800.599	0
73	10377.824	0
74	9181.642	0
75	5376.716	0
76	10652.676	0
77	12158.697	0
78	6566.529	0
79	7945.132	0
80	16656.082	0
81	4050.214	0
82	9745.837	0
83	8249.81	0
84	6810.189	0
85	3305.744	0
86	2282.061	0
87	6904.089	0
88	11210.108	0
89	1028.439	0
90	4196.737	0
91	11715.911	0
92	1817.726	0
93	5673.254	0
94	717.426	0
95	6199.14	0
96	6306.593	0
97	3674.544	0
98	2213.623	0
99	3219.997	0
100	4886.688	0
101	5994.673	0
102	5819.084	0
103	5908.508	0
104	4879.111	0
105	1338.636	0
106	7410.684	0
107	1213.155	0
108	5706.551	0
109	7367.053	0
110	11374.719	0
111	6090.277	0
112	2548.723	0
113	5639.531	0
114	1227.303	0
115	5951.919	0
116	3598.864	0
117	5892.616	0
118	3039.528	0
119	7357.988	0
120	1891.339	0
121	3784.34	0
122	4845.994	0
123	6674.927	0
124	4666.523	0
125	5703.461	0

193	1470.202	0
194	1995.016	0
195	1393.198	0
196	3116.128	0
197	5296.99	0
198	7644.428	0
199	1975.907	0
200	1829.887	0
202	2221.34	0
204	1496.114	0
209	1734.582	0
210	4263.446	0
211	1850.052	0
214	1084.623	0
216	3752.528	0
219	2194.365	0
220	2111.271	0
221	1482.203	0
223	4831.435	0
224	2837.668	0
226	2652.226	0
227	2045.997	0
231	2433.302	0
233	4878.576	0
235	1152.576	0
236	140.24	0
241	3407.205	0
243	2352.224	0
244	1411.586	0
248	3099.849	0
253	5797.433	0
257	1907.972	0
258	2954.507	0
259	1284.858	0
265	1144.189	0
275	2087.888	0
281	3835.723	0
282	1111.012	0
288	2583.46	0
289	1150.588	0
291	4837.203	0
294	3431.214	0
295	10010.406	0
306	2347.36	0
310	1201.517	0
314	1770.364	0
322	5255.569	0
323	1342.32	0
335	2564.477	0
346	1509.282	0
354	7594.412	0
364	7200.342	0
393	5267.32	0
399	3323.9	0
404	1861.308	0
406	8885.498	0
414	1853.155	0
438	2167.973	0
448	1645.077	0
488	4165.946	0
518	7528.718	0
532	4624.189	0
545	2210.97	0
592	7842.921	0
702	2725.864	0
709	8455.579	0
743	7852.387	0

Tabla 45. Longitud en función de Número de Acometidas. Familias

nº acom	Incidencias	Inc./Long	Long. (m)
0	552	3.113502637	177292.286
1	210	3.051466026	68819.38
2	173	2.699122244	64094.911
3	135	2.78942829	48397.014
4	92	2.080495031	44220.245
5	143	3.33249019	42910.854
6	80	2.673758276	29920.431
7	110	3.453945964	31847.632
8	118	3.595439995	32819.349
9	109	3.794387106	28726.642
10	93	3.10946542	29908.678
11	79	2.431542985	32489.658
12	108	3.672490557	29407.836
13	89	2.614440448	34041.701
14	53	2.389844406	22177.176
15	67	3.446577762	19439.573
16	57	3.075122159	18535.849
17	53	3.104652445	17071.154
18	58	2.462643816	23551.924
19	62	2.669884883	23221.975
20	48	2.612202961	18375.295
21	59	3.433833691	17181.962
22	46	3.800827705	12102.627
23	80	4.399699501	18183.06
24	46	2.996828183	15349.562
25	49	2.845366193	17220.982
26	84	3.608973248	23275.318
27	40	2.475674333	16157.214
28	45	3.386374748	13288.547
29	43	2.310164142	18613.396
30	45	7.079963467	6355.965
31	49	3.078351973	15917.608
32	45	3.759365205	11970.106
33	26	3.060240241	8496.065
34	30	4.175019024	7185.596
35	35	5.002210977	6996.906
36	36	3.992660603	9016.544
37	17	2.284966814	7439.933
38	19	2.748407985	6913.093
39	5	0.601061089	8318.622
40	9	1.298293264	6932.178
41	38	3.739655424	10161.364
42	46	6.530284977	7044.103
43	50	8.518273571	5869.734
44	15	1.204243755	12455.95
45	34	2.470922511	13760.043
46	29	4.592336689	6314.868
47	8	0.996719671	8026.329
48	47	5.096840512	9221.399
49	15	1.230168858	12193.448
50	35	4.310474354	8119.756
51	14	1.36620554	10247.36
52	11	0.972690821	11308.835
53	5	0.64154307	7793.709
54	8	1.661649869	4814.492
55	28	4.792078557	5842.976
56	36	8.859479791	4063.444
57	6	1.972174589	3042.327
58	33	9.397069766	3511.733
59	0	0	2940.496
60	31	4.94846651	6264.567
61	5	1.040449557	4805.615
62	20	2.925186027	6837.172
63	0	0	4345.302
64	9	2.586225476	3479.975
65	3	0.399612855	7507.266
66	13	2.347163119	5538.601
67	22	3.013406371	7300.708
68	5	1.117160161	4475.634
69	6	2.010810789	2983.871
70	9	2.030032298	4433.427
71	26	8.322631283	3124.012
72	8	1.186155197	6744.48
73	0	0	3580.902

90	10	4.46254851	2240.872
91	25	3.07950418	8118.19
92	18	9.90248255	1817.726
93	6	2.97268843	2018.375
95	3	3.14106195	955.091
96	10	8.35586632	1196.764
98	4	2.60774553	1533.892
99	1	0.63841306	1566.384
100	33	16.8106781	1963.038
101	0	0	2875.313
102	0	0	1378.876
103	7	2.47529304	2827.948
104	4	1.2572425	3181.566
107	0	0	1213.155
108	0	0	694.087
109	1	1.0475142	954.641
110	31	6.3891969	4851.94
111	9	2.12491928	4235.455
112	1	0.59360568	1684.62
113	9	2.08847766	4309.359
115	0	0	2498.052
116	1	0.41654604	2400.695
117	0	0	1917.835
119	0	0	728.12
120	0	0	476.854
121	0	0	977.246
123	0	0	1546.638
124	0	0	2495.431
125	0	0	1218.862
126	0	0	561.893
128	10	1.42236605	7030.539
129	0	0	600.729
131	4	1.05224723	3801.388
135	9	4.69514428	1916.874
136	0	0	1084.216
137	8	10.977084	728.791
139	0	0	1198.091
140	0	0	1699.86
142	0	0	778.33
143	1	0.50902035	1964.558
148	0	0	1450.634
151	4	5.98803292	667.999
153	10	6.4408629	1552.587
154	29	22.0248927	1316.692
160	5	2.8092034	1779.864
163	27	16.0765386	1679.466
164	55	28.9907019	1897.16
165	15	3.70369639	4050.008
170	1	0.29602168	3378.131
171	0	0	3120.554
174	0	0	1753.169
176	6	2.56962066	2334.975
177	6	1.77121928	3387.497
179	2	2.41738292	827.341
181	0	0	1580.357
182	10	11.6716349	856.778
189	0	0	882.951
194	0	0	1995.016
196	0	0	1433.598
197	2	0.59226198	3376.884
210	1	0.23455205	4263.446
211	2	1.08105069	1850.052
216	0	0	1225.411
223	6	2.88449196	2080.089
231	64	26.3017085	2433.302
235	1	0.86762174	1152.576
243	4	1.70051832	2352.224
248	0	0	3099.849
257	0	0	1907.972
265	0	0	1144.189
275	0	0	2087.888
282	0	0	1111.012
288	0	0	2583.46
289	0	0	1150.588
291	41	8.47597258	4837.203

74	10	1.92690665	5189.665
75	0	0	824.286
76	26	3.633653388	7155.333
77	1	0.192658961	5190.519
78	0	0	1897.114
79	1	0.310142818	3224.321
80	25	5.293713207	4722.583
81	0	0	279.117
82	22	6.025899865	3650.907
83	25	7.279641609	3434.235
84	0	0	4365.699
85	6	4.008851544	1496.688
86	0	0	1247.527
88	9	10.70624329	840.631
89	0	0	1028.439

306	0	0	2347.36
310	14	11.6519367	1201.517
322	37	7.04015112	5255.569
323	0	0	1342.32
335	0	0	2564.477
346	0	0	1509.282
364	1	0.13888229	7200.342
404	0	0	1861.308
414	3	1.61886081	1853.155
448	0	0	1645.077
518	0	0	1699.708
532	53	11.4614692	4624.189
545	20	9.04580343	2210.97
702	10	3.6685616	2725.864
912	0	0	5007.462

Tabla 46. Incidencias acumuladas e Incidencias por unidad de longitud frente a Número de acometidas. Familias.

nº acom	Incidencias	Inc./Long	Long. (m)
0	1430	3.112448507	459445.352
1	465	2.947032547	157785.838
2	376	2.831230749	132804.435
3	329	3.172773996	103694.748
4	343	3.730106989	91954.467
5	290	3.580199708	81001.068
6	183	2.564904452	71347.687
7	148	2.944348081	50265.796
8	182	3.700079674	49188.13
9	106	2.611235491	40593.811
10	92	2.484091822	37035.668
11	118	3.591261159	32857.538
12	92	3.117970692	29506.371
13	78	3.599826987	21667.708
14	56	2.598160595	21553.71
15	50	2.459591494	20328.579
16	56	2.748592389	20374.065
17	29	2.027929059	14300.303
18	35	2.254175264	15526.743
19	21	1.703623258	12326.669
20	30	2.984749423	10051.095
21	20	2.004899775	9975.561
22	11	1.4944375	7360.629
23	15	2.154577738	6961.921
24	9	1.082279063	8315.785
25	5	0.959041082	5213.541
26	11	2.864131338	3840.606
27	12	3.027019937	3964.295
28	24	5.293480483	4533.879
29	5	2.553863536	1957.818
30	1	0.545812201	1832.132
31	3	2.862617272	1047.992
32	7	2.934974795	2385.029
33	11	5.136960713	2141.344
34	2	1.377071891	1452.357
35	0	0	1316.885
36	16	7.491025985	2135.889
37	3	29.80122582	100.667
38	0	0	654.88
39	1	0.8749825	1142.88
40	2	5.998008661	333.444
41	2	1.223662201	1634.438
42	1	3.444036135	290.357
43	0	0	452.378
44	0	0	469.251
45	0	0	305.921
46	0	0	386.015
47	0	0	282.515
48	0	0	141.011
49	0	0	854.908
50	0	0	163

51	0	0	411.041
52	0	0	371.463
54	0	0	164.468
55	0	0	164.702
64	0	0	184.145
65	0	0	334.565
66	0	0	173.627
69	0	0	200.263
70	1	5.659021329	176.709
71	0	0	339.533

Tabla 47. Incidencias acumuladas e incidencias por unidad de longitud frente a Número de Acometidas. Tramos.

Presion Media	RSA
1	82.56
2.5	47.039
2.9	86.68
3.3	46.141
4.1	0.3
4.2	0.996
4.6	21.825
5.1	179.533
5.2	365.128
5.4	172.605
5.5	7.405
5.6	103.026
5.7	20.821
6	0.6
6.5	6.411
6.6	47.167
6.7	239.145
7.2	85.623
7.3	16.417
7.6	287.861
7.7	4.032
7.9	0.3
8	173.911
8.1	155.201
8.3	80.575
8.4	1
8.6	119.949
9.1	70
9.2	2.068
9.3	9.999
9.4	23.751
9.5	342.193
9.8	339.673
10	390.314
10.2	419.249
10.3	166.927
10.4	50.559
10.5	41.319
10.6	489.471
10.7	139.984
11	200.652
11.4	208.623
11.5	414.381
11.6	114.725
11.7	183.833
11.8	59.096
11.9	120.055
12.1	14.191
12.2	522.149
12.3	52.277
12.4	957.504
12.5	509.79
12.6	680.628
12.7	124.992
12.8	239.538
12.9	1.567
13	103.334
13.1	267.801

Presion Media	RSA
27.5	15950.802
27.6	34573.594
27.7	26055.455
27.8	19870.392
27.9	26095.197
28	24591.095
28.1	24533.18
28.2	25054.317
28.3	31504.185
28.4	30999.204
28.5	20534.499
28.6	26646.263
28.7	26403.654
28.8	25583.239
28.9	32629.241
29	19808.034
29.1	27982.266
29.2	26712.485
29.3	37559.755
29.4	30714.759
29.5	42121.134
29.6	28651.47
29.7	35082.717
29.8	37369.828
29.9	35906.358
30	182413.152
30.1	29127.216
30.2	26174.63
30.3	21402.712
30.4	25618.881
30.5	38485.595
30.6	24761.254
30.7	23595.016
30.8	20659.424
30.9	28604.436
31	27353.396
31.1	24053.253
31.2	25076.431
31.3	22474.845
31.4	17944.068
31.5	31111.347
31.6	26643.223
31.7	19812.784
31.8	16281.62
31.9	20895.957
32	15063.755
32.1	24404.179
32.2	14625.999
32.3	18823.966
32.4	17120.592
32.5	11884.932
32.6	12476.148
32.7	7843.061
32.8	12026.15
32.9	13946.139
33	10692.022
33.1	8057.942
33.2	8120.631

Presion Media	RSA
47.6	1398.979
47.7	4678.138
47.8	607.666
47.9	2777.723
48	909.181
48.1	972.237
48.2	2929.677
48.3	395.949
48.4	291.945
48.5	666.957
48.6	5273.333
48.7	1004.85
48.8	750.637
48.9	609.395
49	1336.397
49.1	1582.957
49.2	736.133
49.3	298.981
49.4	2298.54
49.5	1048.733
49.6	2250.573
49.7	1519.053
49.8	698.906
49.9	2143.14
50	284.85
50.1	43.047
50.2	1026.146
50.3	875.432
50.4	171.414
50.5	255.53
50.6	1029.916
50.7	354.52
50.8	51.67
50.9	460.504
51	547.587
51.1	333.897
51.2	352.765
51.3	95.148
51.4	410.737
51.5	2033.459
51.6	93.826
51.7	321.355
51.8	52.634
51.9	133.199
52	130.334
52.1	290.989
52.2	375.875
52.3	295.121
52.4	268.941
52.5	407.239
52.6	7.403
52.7	1087.632
52.9	882.97
53	162.271
53.1	477.494
53.2	1743.719
53.4	583.677
53.5	20.102

13.2	1564.44
13.3	840.821
13.4	1083.689
13.5	0.317
13.6	20.809
13.7	114.896
13.8	665.262
13.9	1.651
14	54.146
14.1	45.685
14.2	525.944
14.3	1949.882
14.4	501.043
14.5	2166.478
14.6	1859.342
14.7	3707.891
14.8	499.108
14.9	2186.257
15	762.611
15.1	820.75
15.2	197.081
15.3	820.22
15.4	955.713
15.5	3715.831
15.6	192.793
15.7	42.424
15.8	733.997
15.9	207.093
16	3996.813
16.1	1209.188
16.2	1638.896
16.3	963.199
16.4	1991.688
16.5	2442.696
16.6	7639.636
16.7	3227.587
16.8	5995.333
16.9	1263.322
17	152.143
17.1	527.136
17.2	3043.695
17.3	1297.933
17.4	1028.965
17.5	2402.528
17.6	1128.191
17.7	3754.58
17.8	966.615
17.9	1740.394
18	1742.538
18.1	1406.319
18.2	1413.751
18.3	1199.289
18.4	11522.568
18.5	1192.579
18.6	2783.742
18.7	2546.718
18.8	2803.603
18.9	1053.04
19	7181.297
19.1	3936.203
19.2	4936.842
19.3	8996.056
19.4	7717.271
19.5	5421.767
19.6	4617.475
19.7	1758.858
19.8	5301.295
19.9	2344.273
20	5283.837
20.1	8215.413
20.2	6150.312
20.3	6718.428
20.4	3924.456
20.5	4556.311
20.6	5079.147
20.7	4131.769
20.8	8614.623
20.9	9045.458

33.3	8564.069
33.4	8129.317
33.5	14344.204
33.6	7086.139
33.7	10389.369
33.8	8746.342
33.9	7391.475
34	4425.388
34.1	6886.961
34.2	3192.636
34.3	4603.959
34.4	8575.122
34.5	9324.314
34.6	2999.963
34.7	8028.595
34.8	3930.28
34.9	5703.434
35	3514.545
35.1	5713.786
35.2	1654.175
35.3	5185.949
35.4	3552.159
35.5	21529.139
35.6	7215.103
35.7	6576.887
35.8	8544.502
35.9	10388.73
36	10753.414
36.1	5493.944
36.2	3560.44
36.3	8763.562
36.4	7349.807
36.5	5961.954
36.6	3928.957
36.7	16936.315
36.8	3203.579
36.9	9404.006
37	3897.506
37.1	3941.984
37.2	2754.061
37.3	5102.202
37.4	3007.597
37.5	9547.022
37.6	2592.622
37.7	7100.485
37.8	9634.469
37.9	7319.694
38	13297.698
38.1	3597.312
38.2	1917.502
38.3	2655.914
38.4	6921.045
38.5	2240.441
38.6	2282.878
38.7	8104.346
38.8	2408.04
38.9	6018.166
39	1755.48
39.1	2987.507
39.2	2869.149
39.3	6056.331
39.4	10930.929
39.5	5395.546
39.6	3545.602
39.7	3339.676
39.8	9840.139
39.9	3407.466
40	4339.89
40.1	2907.248
40.2	4432.616
40.3	3742.631
40.4	5264.51
40.5	802.502
40.6	2143.417
40.7	4967.001
40.8	5152.02
40.9	3303.885
41	2325.187

53.6	1241.291
53.7	96.051
53.8	192.138
53.9	158.625
54	156.166
54.1	2.412
54.2	53.842
54.3	558.295
54.4	21.015
54.5	1201.156
54.6	1177.363
54.7	153.803
54.8	139.076
54.9	19.063
55	2360.87
55.1	435.993
55.2	215.975
55.3	460.089
55.4	55.231
55.5	387.487
55.6	73.688
55.8	2
55.9	103.032
56	1240.082
56.1	94.935
56.2	164.14
56.3	5.714
56.4	18.572
56.5	677.888
56.6	0.3
56.7	521.568
56.8	541.712
56.9	99.364
57	564.009
57.1	2350.348
57.2	29.474
57.3	9.604
57.4	253.02
57.5	319.662
57.6	7.229
57.7	991.125
57.8	295.912
58	39.173
58.1	176.431
58.2	0.5
58.3	11.844
58.4	143.533
58.5	28.734
58.6	240.984
58.7	213.134
58.8	36.483
59	42.187
59.1	78.36
59.3	636.374
59.4	292.084
59.5	26.083
59.6	100.57
59.8	327.345
60	0.501
60.1	234.3
60.2	484.442
60.3	76.352
60.4	27.045
60.5	14.627
60.6	209.715
60.7	170.033
60.9	115.401
61	659.283
61.1	124.933
61.2	770.653
61.7	597.284
61.8	273.759
61.9	4
62.1	0.298
62.2	11.327
62.3	12.988
62.5	258.529
62.6	2.363

21	4769.324	41.1	2516.63	62.7	4.021
21.1	3491.048	41.2	4860.45	62.8	185.51
21.2	12862.352	41.3	6190.248	62.9	0.633
21.3	9570.728	41.4	1209.54	63	1.866
21.4	6829.671	41.5	1141.028	63.1	18.544
21.5	8192.657	41.6	3198.665	63.3	430.431
21.6	6841.776	41.7	1912.775	63.5	459.068
21.7	9056.796	41.8	1164.377	63.7	80.53
21.8	1957.024	41.9	984.776	64.1	92.625
21.9	7067.445	42	3001.895	64.2	1.535
22	9615.441	42.1	1179.009	64.3	80.947
22.1	3810.06	42.2	1860.577	64.4	0.773
22.2	8074.523	42.3	1867.202	64.5	13.473
22.3	9540.231	42.4	1563.811	64.7	1.18
22.4	10420.13	42.5	1423.76	64.9	41.064
22.5	11308.054	42.6	1959.928	65.1	147.433
22.6	11923.813	42.7	5513.925	65.3	246.989
22.7	16314.02	42.8	2795.642	65.5	6.319
22.8	12672.918	42.9	1927.695	65.6	0.688
22.9	9520.644	43	564.379	66	48.09
23	21565.262	43.1	5608.848	66.2	122.166
23.1	10407.578	43.2	2591.263	66.3	0.814
23.2	15579.769	43.3	1607.484	66.4	460.199
23.3	13059.678	43.4	9358.369	66.5	390.648
23.4	19225.451	43.5	2135.968	67.3	0.731
23.5	18184.331	43.6	1450.182	67.8	0.861
23.6	15154.636	43.7	5389.117	67.9	63.917
23.7	5918.347	43.8	3673.576	68	2.472
23.8	7821.045	43.9	356.503	68.1	367.639
23.9	8970.181	44	2617.646	68.2	51.87
24	17357.924	44.1	1914.519	68.3	17.978
24.1	13646.932	44.2	791.859	68.6	308.723
24.2	21008.642	44.3	376.141	69.1	17.196
24.3	11433.639	44.4	4832.656	69.2	62.856
24.4	21185.901	44.5	2463.74	69.3	24.438
24.5	20071.044	44.6	1436.914	69.5	116.689
24.6	9722.051	44.7	2264.884	69.6	0.3
24.7	14236.35	44.8	2371.595	69.8	85.009
24.8	11671.763	44.9	1225.23	69.9	23.205
24.9	15841.72	45	6804.356	70.2	2.999
25	16045.379	45.1	9.263	70.5	19.35
25.1	12306.179	45.2	1225.501	70.6	0.596
25.2	22823.053	45.3	3323.825	71.3	7
25.3	27017.814	45.4	1686.523	72.2	53.235
25.4	21338.519	45.5	3418.57	72.9	2595.831
25.5	16481.73	45.6	1338.739	73.1	152.133
25.6	27846.755	45.7	2263.304	76.3	3.411
25.7	17659.246	45.8	1026.227	76.5	5.156
25.8	16086.822	45.9	4057.321	76.7	240.369
25.9	22233.126	46	860.745	77.4	2.456
26	13981.65	46.1	1063.471	78.3	1087.571
26.1	9814.081	46.2	268.347	87	138.547
26.2	20406.407	46.3	1088.196	98.9	436.157
26.3	30872.263	46.4	1356.352	124.3	1227.3
26.4	21108.804	46.5	1588.442	137	421.264
26.5	22753.975	46.6	3081.237	142	938.808
26.6	34679.785	46.7	129.569	143.4	96.234
26.7	21590.132	46.8	3193.431	145.8	1349.264
26.8	24063.77	46.9	3337.972	149.1	472.291
26.9	25998.942	47	577.783	152.8	0.301
27	19523.175	47.1	267.383	155.6	3.936
27.1	30054.015	47.2	1293.64	155.8	92.624
27.2	27921.274	47.3	5029.644	156.2	424.549
27.3	22576.584	47.4	1094.223	195	196.022
27.4	19162.553	47.5	651.98	365.6	511.786

Tabla 48. Longitud de RSA frente a Presión Media.

Presión Media	RAA
0.2	382.579
1	3.961
1.2	23.023
1.5	4.959
2	45.195
2.1	33.026
2.2	0.5
2.5	181.28
2.6	309.124
2.7	12.889
2.8	15.7
2.9	415.153
3.4	72.698
3.5	164.49
4.3	47.666
4.4	73.564
4.9	2.026
5	432.792
5.1	44.303
5.4	378.437
5.8	1.132
5.9	162.436
6	4.4
6.7	15.276
6.8	273.228
6.9	245.554
7	276.534
7.1	74.651
7.2	1.455
7.7	156.853
7.8	60.128
8.1	326.994
8.7	0.784
8.8	193.083
8.9	179.796
9	0.661
9.1	299.858
9.3	149.67
9.4	756.193
9.5	57.118
9.7	55.698
9.9	267.058
10	0.3
10.5	56.944
10.6	19.999
10.7	31.506
10.9	0.35
11	1.291
11.5	11.609
11.7	1750.821
12	91.086
12.1	576.806
12.2	12.098
12.4	1126.33
12.9	181.992
13.3	11.62
13.7	282.351
13.9	433.731
14.2	1.882
14.4	169.213
14.9	2.881
15	48.124
15.2	77.803
15.3	38.523
15.4	0.758
15.5	11.884

25.2	4335.711
25.3	1678.979
25.4	364.732
25.5	3395.474
25.6	744.864
25.7	2233.469
25.8	2974.687
25.9	1068.268
26	1779.282
26.1	1380.658
26.2	1496.776
26.3	323.702
26.4	756.804
26.5	6754.44
26.6	2897.393
26.7	4069.535
26.8	1466.893
26.9	598.307
27	1875.062
27.1	3131.875
27.2	3124.694
27.3	1934.269
27.4	1892.014
27.5	478.017
27.6	984.606
27.7	1117.834
27.8	3057.532
27.9	2568.962
28	1552.481
28.1	3473.395
28.2	2914.834
28.3	388.434
28.4	971.477
28.5	2063.018
28.6	3207.591
28.7	3405.638
28.8	1425.614
28.9	645.759
29	3908.633
29.1	4428.746
29.2	2912.745
29.3	2825.976
29.4	2104.871
29.5	8270.608
29.6	1042.933
29.7	4232.213
29.8	2499.069
29.9	1878.163
30	24467.182
30.1	2268.852
30.2	1633.987
30.3	1081.16
30.4	2414.769
30.5	3261.463
30.6	925.151
30.7	3193.874
30.8	3139.401
30.9	3878.727
31	1441.415
31.1	11276.589
31.2	5020.238
31.3	3121.547
31.4	1007.561
31.5	4123.513
31.6	1647.624
31.7	1303.762
31.8	238.047

41.2	250.361
41.3	1637.155
41.4	289.96
41.5	257.884
41.6	148.349
42	250.761
42.1	2.624
42.2	156.314
42.4	352.95
42.5	100.613
42.6	1543.593
42.7	330.025
42.8	2657.961
42.9	1021.438
43	3006.446
43.2	817.018
43.3	63.546
43.4	671.162
43.5	234.232
43.6	162.657
43.7	184.839
43.9	447.448
44	511.702
44.1	2523.451
44.2	859.044
44.3	582.025
44.7	441.569
44.8	1337.637
44.9	0.696
45.1	1673.977
45.2	300.653
45.3	501.727
45.4	384.969
45.6	367.972
45.7	635.152
45.8	395.613
45.9	328.582
46	205.473
46.1	0.352
46.2	3.513
46.3	1270.248
46.4	162.86
46.5	701.026
46.6	219.038
46.7	678.906
46.9	988.72
47	35.696
47.1	14.887
47.2	765.893
47.3	984.479
47.4	37.643
47.5	37.921
47.7	247.845
47.8	18.606
47.9	18.894
48	3.433
48.1	1.001
48.2	35.375
48.3	678.588
48.5	400.491
48.6	1685.662
48.7	1950.461
48.8	636
48.9	226.445
49	9
49.1	473.727
49.2	247.483

15.7	24.319
15.8	488.621
15.9	677.746
16	623.532
16.1	19.786
16.3	210.803
16.5	7.985
16.6	93.377
16.7	348.731
16.8	393.386
16.9	97.151
17	1614.868
17.1	3.836
17.3	769.046
17.5	33.418
17.6	32.248
17.7	687.759
17.8	283.381
17.9	144.511
18	0.666
18.1	1006.85
18.2	127.439
18.3	165.749
18.4	29.505
18.5	19.081
18.6	1097.878
18.7	916.332
18.8	659.879
18.9	2988.314
19	399.805
19.1	29.725
19.2	218.205
19.3	1407.5
19.4	636.403
19.5	421.011
19.6	10.288
19.7	612.621
19.8	932.367
20	2.567
20.1	818.621
20.2	2842.365
20.3	644.879
20.4	1118.8
20.5	769.67
20.6	3478.545
20.7	8.48
20.8	2100.671
20.9	492.071
21	293.953
21.1	156.863
21.2	915.176
21.3	767.022
21.4	5034.261
21.5	323.631
21.6	1552.3
21.7	620.882
21.8	207.448
21.9	803.092
22	1677.84
22.1	95.073
22.2	697.178
22.3	2129.884
22.4	655.474
22.5	4312.935
22.6	1337.347
22.7	2.076

31.9	1420.644
32	2095.272
32.1	1090.431
32.2	2017.964
32.3	777.048
32.4	5697.987
32.5	5882.819
32.6	2907.884
32.7	5809.246
32.8	1248.275
32.9	572.311
33	2432.561
33.1	485.504
33.2	2100.251
33.3	623.639
33.4	732.679
33.5	1115.626
33.6	4318.893
33.7	4864.65
33.8	2406.009
33.9	1138.718
34	1977.264
34.1	157.117
34.2	637.931
34.3	445.652
34.4	196.202
34.5	127.78
34.6	631.292
34.7	33.363
34.8	363.902
34.9	271.684
35	1256.931
35.1	871.304
35.2	727.386
35.3	4502.145
35.4	3937.961
35.5	930.794
35.6	81.34
35.7	409.307
35.8	607.438
35.9	2502.204
36	342.281
36.1	838.56
36.2	309.15
36.3	36.042
36.4	1001.637
36.5	137.279
36.6	395.961
36.7	341.262
36.8	2130.386
36.9	265.618
37	714.605
37.1	642.211
37.2	349.353
37.3	111.956
37.4	1.8
37.5	4774.859
37.6	670.23
37.7	6174.609
37.8	2.702
37.9	1239.93
38	808.311
38.1	528.495
38.2	613.099
38.3	672.164
38.4	626.134

49.3	337.464
49.5	4162.128
49.6	11.536
49.7	718.802
49.8	1176.379
49.9	3.632
50.1	6445.471
50.2	961.449
50.3	1305.432
50.4	2.23
50.5	32.452
50.6	144.793
50.9	2245.526
51.1	524.112
51.2	209.929
51.3	0.314
51.4	0.305
51.5	8.744
51.8	657.683
52	416.261
52.5	146.008
52.9	3662.758
53	2.137
53.2	1132.803
53.5	151.288
53.8	242.81
54	1145.64
54.1	354.116
54.2	8.363
54.8	2232.104
54.9	3924.783
55.3	37.321
55.4	1689.263
55.5	739.964
55.6	104.242
55.8	362.873
55.9	1471.219
56	1573.55
56.1	170.057
56.4	1.34
56.5	221.433
56.6	25.466
56.9	1922.451
57.2	392.209
57.5	998.362
57.7	1686.161
57.8	45.342
57.9	4783.095
58	1271.12
58.2	65.611
58.4	941.373
58.5	415.678
58.6	179.736
58.7	415.79
58.8	753.157
58.9	251.239
59	709.789
59.4	115.364
59.8	58.001
60.2	466.303
60.3	333.199
60.4	0.5
60.5	74.292
60.6	1.39
63.5	188.963
64.1	134.246

22.8	4617.433
22.9	534.524
23	1368.515
23.1	2316.904
23.2	5120.256
23.3	3010.683
23.4	1578.796
23.5	90.765
23.6	379.238
23.7	1015.929
23.8	1216.936
23.9	669.949
24	4697.997
24.1	800.569
24.2	2404.784
24.3	2057.693
24.4	176.238
24.5	2212.057
24.6	2366.778
24.7	324.989
24.8	1954.151
24.9	1382.318
25	65.264
25.1	1248.792

38.5	1.851
38.6	456.807
38.7	15.824
38.8	13.15
38.9	218.575
39	18.655
39.1	605.818
39.2	293.351
39.3	289.192
39.4	2.205
39.5	47.95
39.6	1.63
39.7	55.673
39.8	155.513
39.9	338.48
40	381.59
40.1	78.14
40.2	1171.389
40.3	857.257
40.4	1152.802
40.5	1094.545
40.6	277.103
40.8	3031.836
40.9	329.48

64.4	3800.028
64.5	14.035
64.8	649.605
65.9	1.191
66.3	258.863
66.4	689.362
66.6	344.579
66.7	784.817
66.9	684.45
67.2	56.202
67.3	2.511
68.8	648.255
69.1	304.16
69.8	203.462
71.3	26.882
72.7	225.891
75.3	248.526
76.3	11.084
76.4	37.167
76.8	25.941
77.4	4.178
79.9	2
132.7	6685.557

Tabla 49. Longitud de RAA frente a Presión Media.

	Incidencias	Inc./Long	Long. (m)
9.1	0	0	68.244
9.3	0	0	1.829
9.5	0	0	342.193
9.8	0	0	36.776
13.2	0	0	372.701
13.3	0	0	34.854
13.4	0	0	3.001
13.6	0	0	1.085
13.9	0	0	0.6
14	0	0	13.82
14.2	0	0	511.717
14.3	0	0	35.089
14.5	0	0	173.082
14.6	0	0	1267.637
14.9	0	0	21.711
15.1	0	0	0.99
15.4	0	0	11.23
15.6	0	0	0.5
15.8	0	0	2
16.1	2	2.067878099	967.175
16.3	0	0	11.493
16.4	0	0	1094.817
16.5	0	0	836.832
16.6	0	0	54.363
16.7	0	0	2441.017
16.8	0	0	104.586
16.9	0	0	129.455
17.1	1	4.748022449	210.614
17.2	0	0	2648.282
17.3	0	0	6.048
17.4	0	0	142.485
17.5	0	0	986.664
17.6	0	0	71.53
17.7	4	4.780634605	836.709
17.8	0	0	88.549
17.9	0	0	502.483
18	0	0	85.378
18.1	1	0.940512579	1063.25
18.2	5	5.871025316	851.64
18.3	0	0	8.053
18.4	0	0	2042.708
18.5	0	0	247.033
18.6	0	0	285.133
18.7	0	0	365.079
18.8	0	0	997.411

27.9	41	3.12621045	13114.92
28	25	1.72680841	14477.576
28.1	45	3.19643091	14078.202
28.2	14	0.94316082	14843.704
28.3	46	2.25856793	20366.888
28.4	24	1.1954425	20076.248
28.5	37	2.70967998	13654.749
28.6	23	1.44464522	15920.864
28.7	30	1.60203086	18726.231
28.8	44	2.97308527	14799.441
28.9	46	1.97287418	23316.236
29	33	2.30306556	14328.728
29.1	34	2.29076938	14842.175
29.2	65	3.23947301	20064.992
29.3	32	1.24358103	25732.139
29.4	16	1.03978571	15387.786
29.5	73	2.91677953	25027.603
29.6	41	2.04243197	20074.108
29.7	60	2.59530254	23118.692
29.8	59	2.29517624	25706.087
29.9	32	1.26643822	25267.715
30	694	8.11390482	85532.184
30.1	41	2.02048506	20292.157
30.2	44	2.4309305	18100.065
30.3	31	2.06111649	15040.392
30.4	42	2.8466677	14754.093
30.5	48	2.60819783	18403.512
30.6	51	3.38945315	15046.675
30.7	41	2.46117024	16658.742
30.8	26	1.80162626	14431.406
30.9	56	2.72448159	20554.369
31	51	2.63596243	19347.772
31.1	42	3.32185947	12643.521
31.2	44	2.24760483	19576.395
31.3	70	4.62763901	15126.504
31.4	24	1.85319785	12950.587
31.5	56	2.8177826	19873.783
31.6	76	3.62304332	20976.84
31.7	62	4.22472973	14675.495
31.8	57	4.78685077	11907.62
31.9	62	4.58113162	13533.774
32	55	5.56027168	9891.603
32.1	56	3.74659227	14946.916
32.2	25	2.95295633	8466.092
32.3	20	2.68022898	7462.049
32.4	17	1.73293071	9809.971

18.9	1	3.479047437	287.435
19	1	0.603971961	1655.706
19.1	4	1.443658172	2770.739
19.2	1	0.685369016	1459.068
19.3	9	2.186812597	4115.579
19.4	2	1.310478587	1526.16
19.5	3	4.025127529	745.318
19.6	2	0.71741338	2787.793
19.7	0	0	183.892
19.8	1	0.26260718	3807.969
19.9	0	0	502.749
20	1	0.911087018	1097.59
20.1	20	7.235549703	2764.13
20.2	3	2.028197352	1479.146
20.3	5	1.544679229	3236.918
20.4	0	0	1550.353
20.5	4	1.574533495	2540.435
20.6	15	5.131645524	2923.039
20.7	6	3.242829563	1850.236
20.8	4	1.196534716	3342.987
20.9	5	2.175073398	2298.773
21	10	3.850762913	2596.888
21.1	1	1.878636335	532.301
21.2	8	1.939830813	4124.071
21.3	5	1.169026229	4277.064
21.4	9	2.428299074	3706.298
21.5	11	2.330377638	4720.265
21.6	4	1.315782549	3040.016
21.7	1	0.275829959	3625.422
21.8	9	5.828875862	1544.037
21.9	19	4.586449504	4142.638
22	16	5.227874898	3060.517
22.1	14	9.037295628	1549.136
22.2	11	6.123466907	1796.368
22.3	0	0	1590.16
22.4	2	0.61979623	3226.867
22.5	21	3.840714944	5467.732
22.6	38	4.812339649	7896.367
22.7	8	2.020317321	3959.774
22.8	7	1.438347199	4866.697
22.9	5	1.649121661	3031.917
23	5	0.936846257	5337.055
23.1	7	1.667967681	4196.724
23.2	30	3.227139295	9296.159
23.3	13	2.509388975	5180.544
23.4	62	6.260132959	9903.943
23.5	19	2.64721696	7177.349
23.6	12	5.097337915	2354.17
23.7	3	1.48280319	2023.195
23.8	4	1.698390478	2355.171
23.9	36	10.72330478	3357.174
24	17	2.419138861	7027.294
24.1	7	1.459514628	4796.115
24.2	10	1.586955732	6301.373
24.3	13	6.398919469	2031.593
24.4	17	3.715206274	4575.789
24.5	47	5.038971621	9327.3
24.6	8	2.83859769	2818.293
24.7	9	1.660029316	5421.591
24.8	14	3.106625144	4506.498
24.9	15	1.888053055	7944.692
25	20	3.237911541	6176.821
25.1	9	2.547309192	3533.14
25.2	18	2.041077131	8818.873
25.3	28	3.289779009	8511.21
25.4	30	2.8657461	10468.478
25.5	39	4.393712215	8876.321
25.6	35	4.692241887	7459.121
25.7	28	3.132053412	8939.822
25.8	49	10.60061345	4622.374
25.9	82	5.00084252	16397.237
26	43	7.479386029	5749.135
26.1	27	5.265281456	5127.931
26.2	88	6.666240431	13200.844
26.3	121	5.716554518	21166.596

32.5	76	7.72352622	9840.065
32.6	14	2.15886735	6484.882
32.7	15	3.17752919	4720.649
32.8	13	2.11326104	6151.63
32.9	8	1.35487173	5904.618
33	29	3.53290468	8208.543
33.1	7	2.11249392	3313.619
33.2	12	2.59078041	4631.809
33.3	1	1.14917828	870.187
33.4	9	2.45826884	3661.113
33.5	4	3.65755083	1093.628
33.6	0	0	1498.24
33.7	0	0	2851.353
33.8	0	0	599.673
33.9	3	0.86545643	3466.379
34	4	1.45903876	2741.531
34.1	1	1.13561258	880.582
34.2	0	0	974.05
34.3	7	3.14470777	2225.962
34.4	0	0	343.796
34.5	14	3.95276781	3541.822
34.6	1	0.92557798	1080.406
34.7	3	1.2374211	2424.397
34.8	0	0	647.367
34.9	0	0	388.886
35	1	0.97502189	1025.618
35.1	0	0	1585.217
35.2	4	28.4490975	140.602
35.3	3	1.21712692	2464.821
35.4	10	5.95314636	1679.784
35.5	6	2.37704158	2524.146
35.6	1	0.33957391	2944.867
35.7	2	1.6746744	1194.262
35.8	0	0	2349.271
35.9	1	1.26113265	792.938
36	1	0.36319973	2753.306
36.1	0	0	2424.91
36.2	0	0	134.442
36.3	0	0	250.784
36.4	0	0	957.979
36.5	0	0	357.821
36.6	2	3.69440611	541.359
36.7	0	0	5139.223
36.8	0	0	1.335
36.9	0	0	2.458
37	3	2.62699783	1141.988
37.1	0	0	451.75
37.2	0	0	312.828
37.3	0	0	33.058
37.4	0	0	399.842
37.5	0	0	3761.503
37.6	3	7.66036984	391.626
37.7	7	9.80914212	713.62
37.8	0	0	10.426
37.9	3	1.13288174	2648.114
38	6	1.4889154	4029.779
38.1	1	7.26385217	137.668
38.2	0	0	44.989
38.3	2	11.1498879	179.374
38.4	1	0.26503743	3773.052
38.5	4	108.207542	36.966
38.6	0	0	18.514
38.7	0	0	132.837
38.8	0	0	20.201
38.9	0	0	49.081
39.1	4	8.5259669	469.155
39.2	0	0	101.495
39.3	0	0	378.809
39.4	0	0	394.247
39.5	2	1.3869462	1442.017
39.6	1	1.57994856	632.932
39.7	0	0	130.182
39.8	0	0	5.504
39.9	0	0	46.84
40.3	0	0	114.997

26.4	39	4.303589315	9062.203
26.5	68	5.225094778	13014.118
26.6	45	2.653870842	16956.364
26.7	22	2.356895215	9334.314
26.8	60	3.759537005	15959.412
26.9	56	3.021921776	18531.254
27	24	2.145618669	11185.585
27.1	18	0.930401858	19346.479
27.2	25	1.679262366	14887.489
27.3	29	2.106904785	13764.267
27.4	7	0.747765543	9361.223
27.5	25	2.127170597	11752.701
27.6	17	1.00692975	16883.005
27.7	49	3.115213303	15729.26
27.8	23	2.16120402	10642.216

40.6	0	0	0.706
40.9	0	0	6.158
42.2	0	0	212.459
43.4	0	0	108.603
45.7	0	0	0.532
45.9	0	0	11.435
46.3	0	0	29.009
47	0	0	39.261
50.5	0	0	22.405
52.1	0	0	1.974
52.4	0	0	247.879
53.2	0	0	562.8
54	0	0	3
54.3	0	0	0.5
55.1	0	0	113.646
55.6	0	0	5.365
56.6	0	0	0.3
56.8	0	0	332.679
57.5	0	0	199.294
60.6	0	0	45.721
64.9	0	0	41.064
69.6	0	0	0.3
70.2	0	0	2.999

Tabla 50. Incidencias e Incidencias por unidad de longitud frente a Presión Media.

Fluctuación de Presión	RSA
53.3	0.337
51.6	59.016
46.2	2.033
38.5	474.667
33.9	314.615
30.3	26.974
30.2	0.937
26.3	9.515
26.1	348.867
25.9	15.5
24.6	395.408
22.7	6.712
21.8	0.688
21.5	44.363
20.1	2.595
19	6.584
17.5	3.893
16.7	1.065
16.6	1.065
15.7	6.319
15	0.69
14.4	1.535
13.7	8.882
13	29.98
12.4	8823.511
12.3	3559.326
12.2	3608.656
12	11978.982
11.7	2097.876
11.6	228.411
10.9	4.684
10.6	153.789
10.4	164.213
10.3	8.982
10	24.246
9.8	42.04
9.7	189.336
9.5	351.824
9.3	4
9.2	15.946
8.9	1499.245
8.8	42.461
8.7	312.671
8.6	850.169
8.5	0.601
8.4	728.637
8.3	4174.153
8	24.671
7.9	7.026
7.8	56.995

6.3	805.945
6.2	1466.176
6.1	690.492
6	3420.676
5.9	343.21
5.8	853.472
5.7	802.739
5.6	632.877
5.5	23466.564
5.4	6780.806
5.3	15305.449
5.2	10425.916
5.1	13305.763
5	29231.473
4.9	6740.172
4.8	6184.807
4.7	5400.002
4.6	43529.942
4.5	18838.28
4.4	8306.732
4.3	19427.376
4.2	6662.052
4.1	78859.949
4	31734.371
3.9	20925.656
3.8	20163.984
3.7	8457.962
3.6	12894.306
3.5	12965.974
3.4	17640.062
3.3	51274.138
3.2	267750.266
3.1	505737.262
3	196205.367
2.9	86046.116
2.8	14881.134
2.7	11289.807
2.6	28717.322
2.5	188145.188
2.4	15372.028
2.3	65195.138
2.2	33041.012
2.1	34204.791
2	45417.314
1.9	61914.15
1.8	65565.636
1.7	54439.321
1.6	182581.698
1.5	117050.418
1.4	84802.346

7.7	513.723
7.6	1298.9
7.5	4568.898
7.4	1610.219
7.3	38.153
7.2	1390.288
7.1	415.656
7	621.367
6.9	602.209
6.8	286.963
6.7	443.119
6.6	16.474
6.5	145.566
6.4	898.772

1.3	92867.185
1.2	19551.367
1.1	64077.922
1	87997.366
0.9	64597.223
0.8	43087.426
0.7	22204.112
0.6	30634.26
0.5	66715.161
0.4	151065.45
0.3	24840.089
0.2	53027.757
0.1	133281.163
0	37136.699

Tabla 51. Longitud RSA frente a Fluctuación de Presión.

Fluctuación de Presión	RAA
0	9052.515
34.6	605.997
26	292.521
25.9	702.672
25.3	163.239
24.6	4.168
24.4	703.8
23.8	372.685
22.9	2195.68
22.7	7.78
22.1	784.817
21.8	30.581
21.7	317.771
19.1	0.301
17.7	2.511
17.6	3136.348
17.5	699.407
17.3	26.808
17	620.215
16.8	14.035
16.7	391.729
16.6	323.209
15.9	82.004
15.5	3800.028
14.7	59.106
14.5	56.202
14.2	11.62
14	2
13.7	389.94
13.2	258.863
13.1	10.983
12.8	297.633
12.7	1750.821
12.2	104.722
12	121.813
10.9	31.462
10.4	1337.637
10.3	255.765
10.1	984.479
10	894.411
9.7	490.09
9.6	51.83
9.5	530.232
9.2	183.031
8.6	27.465
7.6	0.3
7.5	16.397
7.2	817.018
6.8	0.5
6.4	250.361
6.3	69.405
6.1	627.646
6	241.662
5.8	876.17
5.7	3343.949

5.6	837.893
5.5	0.6
5.4	600.621
5.3	530.054
5.2	279.961
5.1	1.445
5	886.691
4.9	3449.134
4.8	733.858
4.7	1588.948
4.6	4537.608
4.5	1082.706
4.3	7495.428
4.2	3789.112
4.1	11459.95
4	2435.551
3.8	319.318
3.7	1840.501
3.6	413.893
3.5	5326.789
3.4	2165.703
3.3	22891.631
3.2	63507.594
3.1	26735.034
3	5362.424
2.9	3771.359
2.8	2300.225
2.7	1896.344
2.6	4645.171
2.5	18563.702
2.4	1176.444
2.3	4969.572
2.2	7414.836
2.1	2982.468
2	8346.134
1.9	2886.62
1.8	13623.833
1.7	6128.863
1.6	23483.963
1.5	21597.875
1.4	17049.798
1.3	6366.503
1.2	2256.565
1.1	12352.971
1	7497.465
0.9	14450.527
0.8	8026.78
0.7	8587.103
0.6	15782.999
0.5	12684.167
0.4	17419.103
0.3	9762.021
0.2	17560.423
0.1	26520.821

Tabla 52. Longitud de RAA frente a Fluctuación de Presión.

	Incidencias	Inc./Long	Long. (m)
30.3	0	0	26.974
30.2	0	0	0.937
21.5	0	0	44.363
20.1	0	0	2.595
17.5	0	0	3.893
10.9	0	0	1
10.3	0	0	0.743
10	0	0	2.274
9.8	0	0	2.478
8.9	0	0	0.3
8	0	0	24.671
7.9	0	0	4.027
7.6	0	0	447.223
7.1	0	0	13.029
6.9	0	0	6.018
6.8	1	4.88737055	204.609
6.6	0	0	13.467
6.3	0	0	19.379
6.1	0	0	1.999
6	0	0	3420.676
5.9	0	0	250.406
5.8	0	0	104.586
5.7	0	0	678.128
5.6	0	0	564.209
5.5	1	0.11062671	9039.408
5.4	0	0	605.705
5.3	3	0.48411795	6196.837
5.2	1	4.92169582	203.182
5.1	8	2.61349767	3061.032
5	37	1.59032341	23265.708
4.9	1	0.21173355	4722.917
4.8	3	1.67738422	1788.499
4.7	2	1.72543796	1159.126
4.6	0	0	1140.047
4.5	0	0	46.128
4.4	1	1.86853021	535.18
4.3	8	50.2578858	159.179
4.2	0	0	310.585
4.1	3	2.80140556	1070.891
4	33	1.1445583	28832.083

	Incidencias	Inc./Long	Long. (m)
3.9	12	1.03060272	11643.672
3.8	110	6.621335	16612.964
3.7	68	11.1493176	6099.028
3.6	18	1.52009785	11841.343
3.5	12	1.10652365	10844.775
3.4	30	2.16993028	13825.329
3.3	96	2.1674297	44292.094
3.2	804	3.26556176	246205.725
3.1	1449	2.91116799	497738.366
3	661	3.88712753	170048.447
2.9	53	0.96992145	54643.6
2.8	0	0	630.452
2.7	2	4.1443648	482.583
2.6	11	7.80626275	1409.125
2.5	669	9.62719071	69490.677
2.4	0	0	267.206
2.3	2	3.05622538	654.402
2.2	0	0	2064.742
2.1	16	0.76148334	21011.622
2	52	1.90088959	27355.613
1.9	18	2.12608964	8466.247
1.8	81	2.38646812	33941.371
1.7	10	0.67582673	14796.692
1.6	76	1.33815654	56794.551
1.5	10	0.32838223	30452.318
1.4	1	0.20555014	4864.993
1.3	3	0.48710464	6158.841
1.2	8	30.1304649	265.512
1.1	0	0	547.726
1	9	1.18651737	7585.224
0.9	41	4.78501678	8568.413
0.8	8	8.65099611	924.749
0.7	29	8.29791697	3494.853
0.6	2	1.17891024	1696.482
0.5	21	10.1383352	2071.346
0.4	23	8.44274023	2724.234
0.3	14	4.88338477	2866.864
0.2	20	4.45154416	4492.823
0.1	118	1.98266049	59515.989
0	12	4.97483356	2412.141

Tabla 53. Incidencias Acumuladas e Incidencias por unidad de Longitud frente a Fluctuación de Presión.

6.2 Resultados Análisis efecto de filtrado de Outliers. Matrices de Confusión.

6.2.1 Análisis de Bondad por quinquenios con Datos Filtrados. Resultados.

❖ Quinquenio 2003-2008. Resultados para 2009.

Modelo Estadístico para Familias					
				Coincidencias	
Prob. Inc.	Tramos	Long. (Km)	%	Fallos	No Fallos
Bueno: PF<=0,5	2632	194.9	14.9	4	2628
Medio: PF<=0,75	3282	838.5	63.9	102	3180
Alto: PF>0,75	144	37.9	2.9	20	124
Muy Alto: NTR=325	316	240.2	18.3	92	224

		Estimacion		Totales
		Incidencias	No Incid.	
Reales	Incidencias	112	106	218
	No Incid.	348	5808	6156
	Totales	460	5914	

Tabla 54. Resultados por familias para Datos Simplificados. Quinquenio 2003-2008.

❖ Quinquenio 2004-2009. Resultados para 2010.

Modelo Estadístico para Familias					
				Coincidencias	
Prob. Inc.	Tramos	Long. (Km)	%	Fallos	No Fallos
Bueno: PF<=0,5	2989	251.5	19.2	3	2986
Medio: PF<=0,75	3069	839	64	110	2959
Alto: PF>0,75	0	0	0	0	0
Muy Alto: NTR=324	316	221.1	16.9	88	228

		Estimacion		Totales
		Incidencias	No Incid.	
Reales	Incidencias	88	113	201
	No Incid.	228	5945	6173
	Totales	316	6058	

Tabla 55. Resultados por familias para Datos Simplificados. Quinquenio 2004-2009.

❖ Quinquenio 2005-2010. Resultados para 2011.

Modelo Estadístico para Familias					
				Coincidencias	
Prob. Inc.	Tramos	Long. (Km)	%	Fallos	No Fallos
Bueno: PF<=0,5	3200	294	22.4	2	3198
Medio: PF<=0,75	2880	807.3	61.5	93	2787
Alto: PF>0,75	0	0	0	0	0
Muy Alto: NTR=304	294	210.3	16	80	214

		Estimacion		Totales
		Incidencias	No Incid.	
Reales	Incidencias	80	95	175
	No Incid.	214	5985	6199
	Totales	294	6080	

Tabla 56. Resultados por familias para Datos Simplificados. Quinquenio 2005-2010.

❖ **Quinquenio 2006-2011. Resultados para 2012.**

Modelo Estadístico para Familias					
				Coincidencias	
Prob. Inc.	Tramos	Long. (Km)	%	Fallos	No Fallos
Bueno: PF<=0,5	3068	276.6	21.1	6	3062
Medio: PF<=0,75	2977	802.4	61.2	111	2866
Alto: PF>0,75	0	0	0	0	0
Muy Alto: NTR=331	329	232.6	17.7	109	220

Estimacion				
		Incidentes	No Incid.	Totales
Reales	Incidentes	109	117	226
	No Incid.	220	5928	6148
	Totales	329	6045	

Tabla 57. Resultados por familias para Datos Simplificados. Quinquenio 2006-2011.

❖ **Quinquenio 2007-2012. Resultados para 2013.**

Modelo Estadístico para Familias					
				Coincidencias	
Prob. Inc.	Tramos	Long. (Km)	%	Fallos	No Fallos
Bueno: PF<=0,5	3028	264.6	20.2	2	3026
Medio: PF<=0,75	3055	826.1	63	108	2947
Alto: PF>0,75	0	0	0	0	0
Muy Alto: NTR=298	291	220.9	16.8	85	206

Estimacion				
		Incidentes	No Incid.	Totales
Reales	Incidentes	85	110	195
	No Incid.	206	5973	6179
	Totales	291	6083	

Tabla 58. Resultados por familias para Datos Simplificados. Quinquenio 2007-2012.

❖ **Quinquenio 2008-2013. Resultados para 2014.**

Modelo Estadístico para Familias					
				Coincidencias	
Prob. Inc.	Tramos	Long. (Km)	%	Fallos	No Fallos
Bueno: PF<=0,5	3153	281.5	21.5	5	3148
Medio: PF<=0,75	2844	781.5	59.6	105	2739
Alto: PF>0,75	85	33.5	2.6	11	74
Muy Alto: NTR=302	292	215.1	16.4	67	225

Estimacion				
		Incidentes	No Incid.	Totales
Reales	Incidentes	78	110	188
	No Incid.	299	5887	6186
	Totales	377	5997	

Tabla 59. Resultados por familias para Datos Simplificados. Quinquenio 2008-2013.

❖ **Quinquenio 2009-2014. Resultados para 2015.**

Modelo Estadístico para Familias					
				Coincidencias	
Prob. Inc.	Tramos	Long. (Km)	%	Fallos	No Fallos
Bueno: PF<=0,5	3048	260	19.8	9	3039
Medio: PF<=0,75	2880	779.2	59.4	104	2776
Alto: PF>0,75	163	57.2	4.4	24	139
Muy Alto: NTR=297	283	215.2	16.4	77	206

Estimacion				
		Incidencias	No Incid.	Totales
Reales	Incidencias	101	113	214
	No Incid.	345	5815	6160
	Totales	446	5928	

Tabla 60. Resultados por familias para Datos Simplificados. Quinquenio 2009-2014.

❖ **Quinquenio 2010-2015. Resultados para 2016.**

Modelo Estadístico para Familias					
				Coincidencias	
Prob. Inc.	Tramos	Long. (Km)	%	Fallos	No Fallos
Bueno: PF<=0,5	2804	232.3	17.7	9	2795
Medio: PF<=0,75	3214	839	64	138	3076
Alto: PF>0,75	72	27.3	2.1	8	64
Muy Alto: NTR=297	284	213	16.2	64	220

Estimacion				
		Incidencias	No Incid.	Totales
Reales	Incidencias	72	147	219
	No Incid.	284	5871	6155
	Totales	356	6018	

Tabla 61. Resultados por familias para Datos Simplificados. Quinquenio 2010-2015.

❖ **Quinquenio 2011-2016. Resultados para 2017.**

Modelo Estadístico para Familias					
				Coincidencias	
Prob. Inc.	Tramos	Long. (Km)	%	Fallos	No Fallos
Bueno: PF<=0,5	2640	203.1	15.5	7	2633
Medio: PF<=0,75	3445	900.4	68.6	146	3299
Alto: PF>0,75	0	0	0	0	0
Muy Alto: NTR=304	289	208.1	15.9	82	207

Estimacion				
		Incidencias	No Incid.	Totales
Reales	Incidencias	82	153	235
	No Incid.	207	5932	6139
	Totales	289	6085	

Tabla 62. Resultados por familias para Datos Simplificados. Quinquenio 2011-2016.

❖ Quinquenio 2012-2017. Resultados para 2018.

Modelo Estadístico para Familias					
				Coincidencias	
Prob. Inc.	Tramos	Long. (Km)	%	Fallos	No Fallos
Bueno: PF<=0,5	2685	205.1	15.6	6	2679
Medio: PF<=0,75	3363	884.2	67.4	140	3223
Alto: PF>0,75	20	10	0.8	4	16
Muy Alto: NTR=321	306	212.3	16.2	88	218

Estimacion				
		Incidencias	No Incid.	Totales
Reales	Incidencias	92	146	238
	No Incid.	234	5902	6136
	Totales	326	6048	

Tabla 63. Resultados por familias para Datos Simplificados. Quinquenio 2012-2017.

6.2.2 Análisis de Bondad por quinquenios con Datos Originales. Resultados.

❖ Quinquenio 2003-2008. Resultados para 2009.

Modelo Estadístico para Familias					
				Coincidencias	
Prob. Inc.	Tramos	Long. (Km)	%	Fallos	No Fallos
Bueno: PF<=0,5	3287	266.9	18.5	27	3260
Medio: PF<=0,75	3405	915.8	63.6	102	3303
Alto: PF>0,75	0	0	0	0	0
Muy Alto: NTR=281	281	256.6	17.8	105	176

Estimacion				
		Incidencias	No Incid.	Totales
Reales	Incidencias	105	129	234
	No Incid.	176	6563	6739
	Totales	281	6692	

Tabla 64. Resultados por familias para Datos Originales. Quinquenio 2003-2008.

❖ Quinquenio 2004-2009. Resultados para 2010.

Modelo Estadístico para Familias					
				Coincidencias	
Prob. Inc.	Tramos	Long. (Km)	%	Fallos	No Fallos
Bueno: PF<=0,5	3451	282	19.6	24	3427
Medio: PF<=0,75	3196	831.3	57.8	97	3099
Alto: PF>0,75	0	0	0	0	0
Muy Alto: NTR=326	326	326	22.6	90	236

Estimacion				
		Incidencias	No Incid.	Totales
Reales	Incidencias	90	121	211
	No Incid.	236	6526	6762
	Totales	326	6647	

Tabla 65. Resultados por familias para Datos Originales. Quinquenio 2004-2009.

❖ **Quinquenio 2005-2010. Resultados para 2011.**

Modelo Estadístico para Familias					
				Coincidencias	
Prob. Inc.	Tramos	Long. (Km)	%	Fallos	No Fallos
Bueno: PF<=0,5	3648	309.8	21.5	18	3630
Medio: PF<=0,75	3018	796.2	55.3	84	2934
Alto: PF>0,75	0	0	0	0	0
Muy Alto: NTR=307	307	333.3	23.2	84	223

Estimacion				
		Incidentes	No Incid.	Totales
Reales	Incidentes	84	102	186
	No Incid.	223	6564	6787
	Totales	307	6666	

Tabla 66. Resultados por familias para Datos Originales. Quinquenio 2005-2010.

❖ **Quinquenio 2006-2011. Resultados para 2012.**

Modelo Estadístico para Familias					
				Coincidencias	
Prob. Inc.	Tramos	Long. (Km)	%	Fallos	No Fallos
Bueno: PF<=0,5	3536	302.5	21	26	3510
Medio: PF<=0,75	3104	781.1	54.3	104	3000
Alto: PF>0,75	0	0	0	0	0
Muy Alto: NTR=333	333	355.7	24.7	104	229

Estimacion				
		Incidentes	No Incid.	Totales
Reales	Incidentes	104	130	234
	No Incid.	229	6510	6739
	Totales	333	6640	

Tabla 67. Resultados por familias para Datos Originales. Quinquenio 2006-2011.

❖ **Quinquenio 2007-2012. Resultados para 2013.**

Modelo Estadístico para Familias					
				Coincidencias	
Prob. Inc.	Tramos	Long. (Km)	%	Fallos	No Fallos
Bueno: PF<=0,5	3419	274.5	19.1	15	3404
Medio: PF<=0,75	3254	861.6	59.9	95	3159
Alto: PF>0,75	0	0	0	0	0
Muy Alto: NTR=300	300	303.1	21.1	94	206

Estimacion				
		Incidentes	No Incid.	Totales
Reales	Incidentes	94	110	204
	No Incid.	206	6563	6769
	Totales	300	6673	

Tabla 68. Resultados por familias para Datos Originales. Quinquenio 2007-2012.

❖ Quinquenio 2008-2013. Resultados para 2014.

Modelo Estadístico para Familias					
				Coincidencias	
Prob. Inc.	Tramos	Long. (Km)	%	Fallos	No Fallos
Bueno: PF<=0,5	3488	281.8	19.6	20	3468
Medio: PF<=0,75	3182	868.2	60.3	102	3080
Alto: PF>0,75	0	0	0	0	0
Muy Alto: NTR=303	303	289.3	20.1	76	227

Estimacion				
		Incidencias	No Incid.	Totales
Reales	Incidencias	76	122	198
	No Incid.	227	6548	6775
	Totales	303	6670	

Tabla 69. Resultados por familias para Datos Originales. Quinquenio 2008-2013.

❖ Quinquenio 2009-2014. Resultados para 2015.

Modelo Estadístico para Familias					
				Coincidencias	
Prob. Inc.	Tramos	Long. (Km)	%	Fallos	No Fallos
Bueno: PF<=0,5	3396	269.1	18.7	26	3370
Medio: PF<=0,75	3279	885.5	61.5	112	3167
Alto: PF>0,75	0	0	0	0	0
Muy Alto: NTR=298	298	284.6	19.8	85	213

Estimacion				
		Incidencias	No Incid.	Totales
Reales	Incidencias	85	138	223
	No Incid.	213	6537	6750
	Totales	298	6675	

Tabla 70. Resultados por familias para Datos Originales. Quinquenio 2009-2014.

❖ Quinquenio 2010-2015. Resultados para 2016.

Modelo Estadístico para Familias					
				Coincidencias	
Prob. Inc.	Tramos	Long. (Km)	%	Fallos	No Fallos
Bueno: PF<=0,5	3126	233.3	16.2	22	3104
Medio: PF<=0,75	3548	920.7	64	129	3419
Alto: PF>0,75	0	0	0	0	0
Muy Alto: NTR=299	299	285.2	19.8	76	223

Estimacion				
		Incidencias	No Incid.	Totales
Reales	Incidencias	76	151	227
	No Incid.	223	6523	6746
	Totales	299	6674	

Tabla 71. Resultados por familias para Datos Originales. Quinquenio 2010-2015.

❖ **Quinquenio 2011-2016. Resultados para 2017.**

Modelo Estadístico para Familias					
				Coincidencias	
Prob. Inc.	Tramos	Long. (Km)	%	Fallos	No Fallos
Bueno: PF<=0,5	3126	233.3	16.2	22	3104
Medio: PF<=0,75	3548	920.7	64	129	3419
Alto: PF>0,75	0	0	0	0	0
Muy Alto: NTR=299	299	285.2	19.8	76	223

		Estimacion		
		Incidencias	No Incid.	Totales
Reales	Incidencias	85	159	244
	No Incid.	221	6508	6729
	Totales	306	6667	

Tabla 72. Resultados por familias para Datos Originales. Quinquenio 2011-2016.

❖ **Quinquenio 2012-2017. Resultados para 2018.**

Modelo Estadístico para Familias					
				Coincidencias	
Prob. Inc.	Tramos	Long. (Km)	%	Fallos	No Fallos
Bueno: PF<=0,5	3105	235.6	16.4	23	3082
Medio: PF<=0,75	3546	888.1	61.7	132	3414
Alto: PF>0,75	0	0	0	0	0
Muy Alto: NTR=322	322	315.6	21.9	91	231

		Estimacion		
		Incidencias	No Incid.	Totales
Reales	Incidencias	91	155	246
	No Incid.	231	6496	6727
	Totales	322	6651	

Tabla 73. Resultados por familias para Datos Originales. Quinquenio 2012-2017.

6.3 Resultados Análisis de Factores Cuantitativos. Matrices de Confusión.

6.3.1 Análisis de bondad por quinquenios. Datos normalizados.

En las tablas 43-52 se muestran los resultados obtenidos mediante el análisis de bondad por quinquenios con el uso de datos normalizados.

❖ Quinquenio 2003-2008. Predicción 2009

Modelo Estadístico para Familias					
				Coincidencias	
Prob. Inc.	Tramos	Long. (Km)	%	Fallos	No Fallos
Buena: PF<=0,5	3241	381.1	26.5	12	3229
Medio: PF<=0,75	3406	799.3	55.5	117	3289
Alto: PF>0,75	0	0	0	0	0
Muy Alto: NTR=326	326	258.9	18	96	230

		Estimacion		Totales
		Incidentes	No Incid.	
Reales	Incidentes	96	129	225
	No Incid.	230	6518	6748
	Totales	326	6647	

Tabla 74. Resultados análisis de bondad datos originales. Quinquenio 2003-2008.

❖ Quinquenio 2004-2009. Predicción 2010

Modelo Estadístico para Familias					
				Coincidencias	
Prob. Inc.	Tramos	Long. (Km)	%	Fallos	No Fallos
Buena: PF<=0,5	3721	486	33.8	33	3688
Medio: PF<=0,75	2926	685.5	47.6	93	2833
Alto: PF>0,75	0	0	0	0	0
Muy Alto: NTR=326	326	267.7	18.6	85	241

		Estimacion		Totales
		Incidentes	No Incid.	
Reales	Incidentes	85	126	211
	No Incid.	241	6521	6762
	Totales	326	6647	

Tabla 75. Resultados análisis de bondad datos originales. Quinquenio 2004-2009.

❖ Quinquenio 2005-2010. Predicción 2011

Modelo Estadístico para Familias					
				Coincidencias	
Prob. Inc.	Tramos	Long. (Km)	%	Fallos	No Fallos
Buena: PF<=0,5	3873	510.3	35.5	16	3857
Medio: PF<=0,75	2793	663.9	46.1	90	2703
Alto: PF>0,75	0	0	0	0	0
Muy Alto: NTR=307	307	265.1	18.4	80	227

		Estimacion		Totales
		Incidentes	No Incid.	
Reales	Incidentes	80	106	186
	No Incid.	227	6560	6787
	Totales	307	6666	

Tabla 76. Resultados análisis de bondad datos originales. Quinquenio 2005-2010.

❖ **Quinquenio 2006-2011. Predicción 2012**

Modelo Estadístico para Familias				Coincidencias	
Prob. Inc.	Tramos	Long. (Km)	%	Fallos	No Fallos
Bueno: PF<=0,5	3817	520	36.1	28	3789
Medio: PF<=0,75	2823	635.6	44.2	100	2723
Alto: PF>0,75	0	0	0	0	0
Muy Alto: NTR=333	333	283.7	19.7	106	227

		Estimacion		Totales
		Incidencias	No Incid.	
Reales	Incidencias	106	128	234
	No Incid.	227	6512	6739
	Totales	333	6640	

Tabla 77. Resultados análisis de bondad datos originales. Quinquenio 2006-2011.

❖ **Quinquenio 2007-2012. Predicción 2013**

Modelo Estadístico para Familias				Coincidencias	
Prob. Inc.	Tramos	Long. (Km)	%	Fallos	No Fallos
Bueno: PF<=0,5	3671	484.1	33.6	17	3654
Medio: PF<=0,75	3002	705.9	49	88	2914
Alto: PF>0,75	0	0	0	0	0
Muy Alto: NTR=300	300	249.3	17.3	99	201

		Estimacion		Totales
		Incidencias	No Incid.	
Reales	Incidencias	99	105	204
	No Incid.	201	6568	6769
	Totales	300	6673	

Tabla 78. Resultados análisis de bondad datos originales. Quinquenio 2007-2012.

❖ **Quinquenio 2008-2013. Predicción 2014**

Modelo Estadístico para Familias				Coincidencias	
Prob. Inc.	Tramos	Long. (Km)	%	Fallos	No Fallos
Bueno: PF<=0,5	3696	483	33.6	22	3674
Medio: PF<=0,75	2974	720.9	50.1	104	2870
Alto: PF>0,75	0	0	0	0	0
Muy Alto: NTR=303	303	235.4	16.4	72	231

		Estimacion		Totales
		Incidencias	No Incid.	
Reales	Incidencias	72	126	198
	No Incid.	231	6544	6775
	Totales	303	6670	

Tabla 79. Resultados análisis de bondad datos originales. Quinquenio 2008-2013.

❖ **Quinquenio 2009-2014. Predicción 2015**

Modelo Estadístico para Familias					
				Coincidencias	
Prob. Inc.	Tramos	Long. (Km)	%	Fallos	No Fallos
Bueno: PF<=0,5	3650	462.9	32.2	27	3623
Medio: PF<=0,75	3025	739.9	51.4	112	2913
Alto: PF>0,75	0	0	0	0	0
Muy Alto: NTR=298	298	236.5	16.4	84	214

		Estimacion		Totales
		Incidencias	No Incid.	
Reales	Incidencias	84	139	223
	No Incid.	214	6536	6750
	Totales	298	6675	

Tabla 80. Resultados análisis de bondad datos originales. Quinquenio 2009-2014.

❖ **Quinquenio 2010-2015. Predicción 2016**

Modelo Estadístico para Familias					
				Coincidencias	
Prob. Inc.	Tramos	Long. (Km)	%	Fallos	No Fallos
Bueno: PF<=0,5	3362	426.2	29.6	23	3339
Medio: PF<=0,75	3312	771	53.6	129	3183
Alto: PF>0,75	0	0	0	0	0
Muy Alto: NTR=299	299	242.1	16.8	75	224

		Estimacion		Totales
		Incidencias	No Incid.	
Reales	Incidencias	75	152	227
	No Incid.	224	6522	6746
	Totales	299	6674	

Tabla 81. Resultados análisis de bondad datos originales. Quinquenio 2010-2015.

❖ **Quinquenio 2011-2016. Predicción 2017**

Modelo Estadístico para Familias					
				Coincidencias	
Prob. Inc.	Tramos	Long. (Km)	%	Fallos	No Fallos
Bueno: PF<=0,5	3361	432.2	30	42	3319
Medio: PF<=0,75	3306	743.7	51.7	118	3188
Alto: PF>0,75	0	0	0	0	0
Muy Alto: NTR=306	306	263.5	18.3	84	222

		Estimacion		Totales
		Incidencias	No Incid.	
Reales	Incidencias	84	160	244
	No Incid.	222	6507	6729
	Totales	306	6667	

Tabla 82. Resultados análisis de bondad datos originales. Quinquenio 2011-2016.

❖ **Quinquenio 2012-2017. Predicción 2018**

Modelo Estadístico para Familias					
				Coincidencias	
Prob. Inc.	Tramos	Long. (Km)	%	Fallos	No Fallos
Bueno: PF<=0,5	3334	417.9	29	31	3303
Medio: PF<=0,75	3317	764.2	53.1	122	3195
Alto: PF>0,75	0	0	0	0	0
Muy Alto: NTR=322	322	257.1	17.9	93	229

		Estimacion		Totales
		Incidencias	No Incid.	
Reales	Incidencias	93	153	246
	No Incid.	229	6498	6727
	Totales	322	6651	

Tabla 83. Resultados análisis de bondad datos originales. Quinquenio 2012-2017.

6.3.2 Resultados de análisis de bondad para datos normalizados.

En las tablas 53-62 se muestran los resultados obtenidos mediante el análisis de bondad por quinquenios con el uso de datos normalizados.

❖ Quinquenio 2003-2008. Predicción 2009

Modelo Estadístico para Familias					
				Coincidencias	
Prob. Inc.	Tramos	Long. (Km)	%	Fallos	No Fallos
Bueno: PF<=0,5	3900	757.7	48.8	36	3864
Medio: PF<=0,75	3620	595.4	38.3	108	3512
Alto: PF>0,75	0	0	0	0	0
Muy Alto: NTR=305	305	199.7	12.9	97	208

		Estimacion		Totales
		Incidencias	No Incid.	
Reales	Incidencias	97	144	241
	No Incid.	208	7376	7584
	Totales	305	7520	

Tabla 84. Resultados análisis de bondad datos normalizados. Quinquenio 2003-2008.

❖ Quinquenio 2004-2009. Predicción 2010

Modelo Estadístico para Familias					
				Coincidencias	
Prob. Inc.	Tramos	Long. (Km)	%	Fallos	No Fallos
Bueno: PF<=0,5	4553	890.1	57.3	48	4505
Medio: PF<=0,75	2969	474.2	30.5	97	2872
Alto: PF>0,75	0	0	0	0	0
Muy Alto: NTR=303	303	188.5	12.1	82	221

		Estimacion		Totales
		Incidencias	No Incid.	
Reales	Incidencias	82	145	227
	No Incid.	221	7377	7598
	Totales	303	7522	

Tabla 85. Resultados análisis de bondad datos normalizados. Quinquenio 2004-2009.

❖ **Quinquenio 2005-2010. Predicción 2011**

Modelo Estadístico para Familias					
				Coincidencias	
Prob. Inc.	Tramos	Long. (Km)	%	Fallos	No Fallos
Bueno: PF<=0,5	4757	928.1	59.8	31	4726
Medio: PF<=0,75	2781	433.5	27.9	95	2686
Alto: PF>0,75	0	0	0	0	0
Muy Alto: NTR=287	287	191.2	12.3	81	206

Estimacion				
		Incidencias	No Incid.	Totales
Reales	Incidencias	81	126	207
	No Incid.	206	7412	7618
	Totales	287	7538	

Tabla 86. Resultados análisis de bondad datos normalizados. *Quinquenio 2005-2010.*

❖ **Quinquenio 2006-2011. Predicción 2012**

Modelo Estadístico para Familias					
				Coincidencias	
Prob. Inc.	Tramos	Long. (Km)	%	Fallos	No Fallos
Bueno: PF<=0,5	4508	901.3	58	46	4462
Medio: PF<=0,75	3005	451.5	29.1	97	2908
Alto: PF>0,75	0	0	0	0	0
Muy Alto: NTR=312	312	200	12.9	106	206

Estimacion				
		Incidencias	No Incid.	Totales
Reales	Incidencias	106	143	249
	No Incid.	206	7370	7576
	Totales	312	7513	

Tabla 87. Resultados análisis de bondad datos normalizados. *Quinquenio 2006-2011.*

❖ **Quinquenio 2007-2012. Predicción 2013**

Modelo Estadístico para Familias					
				Coincidencias	
Prob. Inc.	Tramos	Long. (Km)	%	Fallos	No Fallos
Bueno: PF<=0,5	4448	865.7	55.7	32	4416
Medio: PF<=0,75	3094	500.1	32.2	95	2999
Alto: PF>0,75	0	0	0	0	0
Muy Alto: NTR=283	283	187	12	99	184

Estimacion				
		Incidencias	No Incid.	Totales
Reales	Incidencias	99	127	226
	No Incid.	184	7415	7599
	Totales	283	7542	

Tabla 88. Resultados análisis de bondad datos normalizados. *Quinquenio 2007-2012.*

❖ **Quinquenio 2008-2013. Predicción 2014**

Modelo Estadístico para Familias					
Prob. Inc.	Tramos	Long. (Km)	%	Coincidencias	
				Fallos	No Fallos
Bueno: PF<=0,5	4479	866.6	55.8	33	4446
Medio: PF<=0,75	3066	495.8	31.9	108	2958
Alto: PF>0,75	0	0	0	0	0
Muy Alto: NTR=280	280	190.4	12.3	80	200

		Estimacion		Totales
		Incidentes	No Incid.	
Reales	Incidentes	80	141	221
	No Incid.	200	7404	7604
	Totales	280	7545	

Tabla 89. Resultados análisis de bondad datos normalizados. Quinquenio 2008-2013.

❖ Quinquenio 2009-2014. Predicción 2015

Modelo Estadístico para Familias					
Prob. Inc.	Tramos	Long. (Km)	%	Coincidencias	
				Fallos	No Fallos
Bueno: PF<=0,5	4503	887.1	57.1	39	4464
Medio: PF<=0,75	3049	483	31.1	122	2927
Alto: PF>0,75	0	0	0	0	0
Muy Alto: NTR=273	273	182.7	11.8	81	192

		Estimacion		Totales
		Incidentes	No Incid.	
Reales	Incidentes	81	161	242
	No Incid.	192	7391	7583
	Totales	273	7552	

Tabla 90. Resultados análisis de bondad datos normalizados. Quinquenio 2009-2014.

❖ Quinquenio 2010-2015. Predicción 2016

Modelo Estadístico para Familias					
Prob. Inc.	Tramos	Long. (Km)	%	Coincidencias	
				Fallos	No Fallos
Bueno: PF<=0,5	4230	841	54.2	49	4181
Medio: PF<=0,75	3323	528.1	34	120	3203
Alto: PF>0,75	0	0	0	0	0
Muy Alto: NTR=272	272	183.8	11.8	73	199

		Estimacion		Totales
		Incidentes	No Incid.	
Reales	Incidentes	73	169	242
	No Incid.	199	7384	7583
	Totales	272	7553	

Tabla 91. Resultados análisis de bondad datos normalizados. Quinquenio 2010-2015.

❖ **Quinquenio 2011-2016. Predicción 2017**

Modelo Estadístico para Familias				Coincidencias	
Prob. Inc.	Tramos	Long. (Km)	%	Fallos	No Fallos
Bueno: PF<=0,5	4102	855.9	55.1	64	4038
Medio: PF<=0,75	3449	518.1	33.4	120	3329
Alto: PF>0,75	0	0	0	0	0
Muy Alto: NTR=274	274	178.8	11.5	79	195

		Estimacion		
		Incidentes	No Incid.	Totales
Reales	Incidentes	79	184	263
	No Incid.	195	7367	7562
	Totales	274	7551	

Tabla 92. Resultados análisis de bondad datos normalizados. Quinquenio 2011-2016.

❖ **Quinquenio 2012-2017. Predicción 2018**

Modelo Estadístico para Familias				Coincidencias	
Prob. Inc.	Tramos	Long. (Km)	%	Fallos	No Fallos
Bueno: PF<=0,5	3921	791.8	51	33	3888
Medio: PF<=0,75	3612	580.5	37.4	138	3474
Alto: PF>0,75	0	0	0	0	0
Muy Alto: NTR=292	292	180.5	11.6	90	202

		Estimacion		
		Incidentes	No Incid.	Totales
Reales	Incidentes	90	171	261
	No Incid.	202	7362	7564
	Totales	292	7533	

Tabla 93. Resultados análisis de bondad datos normalizados. Quinquenio 2012-2017.