

TRABAJO FIN DE MÁSTER

ANÁLISIS DE LA FIABILIDAD DE UN SISTEMA DE TELECONTROL
FRENTE A FALLOS ACCIDENTALES O PROVOCADOS. APLICACIÓN AL
SISTEMA DE ABASTECIMIENTO DE C-TOWN

INTENSIFICACIÓN: HIDRÁULICA URBANA

Autor:

BRAYAN JULIAN TORRES CHAZATAR

Director:

JAVIER MARTINEZ SOLANO

Codirector:

PEDRO L. IGLESIAS REY

SEPTIEMBRE, 2017



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA

máster en ingeniería
hidráulica y medio ambiente
mihma



Resumen del Trabajo Fin de Máster

Datos del proyecto

Título del TFM en español: ANALISIS DE LA FIABILIDAD DE UN SISTEMA DE TELECONTROL FRENTE A FALLOS ACCIDENTALES O PROVOCADOS. APLICACIÓN AL SISTEMA DE ABASTECIMIENTO DE C-TOWN.

Título del TFM en inglés: ANALYSIS OF THE RELIABILITY OF A TELECONTROL SYSTEM IN RESPONSE TO ACCIDENTAL OR CONSEQUENTIAL FAILURES. APPLICATION TO THE C-TOWN SUPPLY SYSTEM

Título del TFM en Valenciano: ANALISIS DE LA FIABILIDAD D'UN SISTEMA DE TELECONTROL FRONT A FALLOS ACCIDENTALS O PROVOCATS. APLICACIO AL SISTEMA D'ABASTIMENT DE C-TOWN.

Alumno: BRAYAN JULIAN TORRES CHAZATAR

Director: JAVIER MARTINEZ SOLANO

Codirector: PEDRO LUIS IGLESIAS REY

Fecha de Lectura: Septiembre, 2017

RESUMEN

En español: La comprensión, el análisis y el mejoramiento de los procesos relacionados con los abastecimientos urbanos se ha convertido en uno de los principales objetivos de la mayoría de países a nivel mundial. Uno de los principales intereses de estos mismos radica en el buen aprovechamiento del recurso que cada vez es más escaso, así como la protección de la infraestructura y un buen control del consumo de los abonados. El empleo de la telemetría y el telecontrol de las redes facilita la operación de las mismas, dando respuestas eficaces a cambios en las circunstancias de funcionamiento.

Hoy en día y sobre todo a raíz del resurgimiento de determinados grupos terroristas de carácter global ha crecido la preocupación por la seguridad frente a ataques. Una de las muchas medidas tomadas a tal efecto fue la aprobación de diversas normativas en las que se establecen medidas para la protección de las infraestructuras críticas. Todas estas normativas coinciden en incluir las infraestructuras hidráulicas (tanto de agua potable como de aguas residuales) en el catálogo de infraestructuras críticas.

De entre los muchos puntos vulnerables que se pueden encontrar en las redes de distribución de agua, los sistemas de telecontrol tienen como riesgo añadido la posibilidad de sufrir no solo fallos físicos, sino también ataques cibernéticos. Por todo ello, cualquier estrategia que permita establecer un sistema de control de la calidad de las señales recibidas supone una importante aportación al problema. Este trabajo pretende establecer una estrategia de control de las señales recibidas por el sistema de telecontrol de forma que ante un fallo fortuito o intencionado, éste sea detectado y corregido a tiempo.

En el caso de la red C-Town, es imprescindible obtener un análisis que permita conocer el estado real del sistema, debido a que en los últimos años se logró implementar una tecnología novedosa para la recogida de datos a distancia a través del proceso que llevan a cabo los sensores de campos respectivos y el control remoto de actuadores. Sin embargo, después de que esta tecnología fuese introducida se empezó a detectar niveles anómalos en los tanques, por ejemplo, rebosamiento en algunos de los tanques de forma inesperada mientras que las lecturas del nivel de agua siempre estuvieron por debajo de los umbrales de alarma y las maniobras de bombeo también parecían ser normales.

Es por eso que este proyecto busca brindar protección a la red de abastecimiento C-TOWN, creando así un entorno de certidumbre en cuanto al comportamiento del sistema que depende en gran parte de las medidas tomadas por los aparatos correspondientes, en este caso se cuenta con un registro hora a hora de algunas variables, las cuales se usaran para crear un modelo que pueda detectar inconsistencias en puntos críticos de la red, ya sean de carácter accidental o provocados.

La red proporciona un conjunto de datos de telemedida previos a las averías, es decir, los valores antes mencionados sin ningún tipo de manipulación, fallo o accidente en un periodo de 1 año. Estos datos describen un comportamiento normal del sistema y por ende son los que se usarán para la creación del modelo de detección. Además, se proporcionan datos anómalos donde se registran de una manera incorrecta valores durante un periodo de tiempo igual al anterior. Precisamente la comparación del comportamiento de estos conjuntos de datos dará como resultado una metodología para la detección de anomalías en los datos proporcionados por la telemedida, por lo cual seremos capaces de determinar rangos de comportamiento verdaderos de los elementos hidráulicos y así mismos comportamientos no inusuales de los mismos que alerten sobre un posible fallo.

Para llevar a cabo esta propuesta será necesario construir un modelo matemático donde sea posible ingresar los datos registrados y este pueda determinar si son correctos o no, también será necesario estimar las demandas y por ende los patrones de consumos en cada periodo de tiempo, debido a que la red carece de datos de facturación por lo tanto es imprescindible saber cuál es el suministro verdadero en el instante respectivo.

Por otro lado, será necesario hacer pruebas de calidad de las mediciones obtenidas en la red para estar seguros que la información suministrada es correcta y que puede ser procesada como fundamento de la comprensión real del sistema. Esto nos lleva así mismo a plantear no solo pruebas de los datos verdaderos, sino que posteriormente a esto, evaluar el conjunto de datos que contiene fallos conocidos o provocados debido a que se planteara entornos de fallos simulados, usando el programa EPANET.

Por último, se recogerá y se evaluará la información obtenida a través de una matriz de confusión, la cual contiene una tabla de operación binaria de doble entrada. Se contrastará por una parte la respuesta del sistema de detección (situación segura o situación anómala) con el funcionamiento efectivo del sistema de telecontrol. Si ambas respuestas coinciden se tendrán aciertos (positivo si hay fallo y negativo si no lo hay). En caso de que haya discrepancias entre la predicción del sistema de detección y el funcionamiento de la red se tendrá un fallo del sistema (falso positivo si el sistema avisa y no hay fallo o falso negativo si hay fallo y no ha sido avisado). Estas discrepancias servirán para medir la fiabilidad del sistema de detección.

Las columnas de la matriz representan las instancias de una clase real mientras que las filas representan las instancias de una clase predicha. Esta matriz dará como resultado a su vez dos estados del sistema, estado de FALLO O SEGURO, lo cual garantizará claridad acerca del comportamiento del sistema cuando se presente cualquier tipo de anomalía descritas anteriormente.



ABSTRACT

En ingles: Understanding, analyzing and improving processes related to urban supplies has become one of the main objectives of most countries worldwide. One of the main interests of these is the good use of the resource that is increasingly scarce, as well as protection of infrastructure and good control of subscribers' consumption. The use of telemetry and remote control of the networks facilitates the operation of the same, giving effective responses to changes in operating circumstances.

Today, and especially as a result of the resurgence of certain terrorist groups of a global character, the concern for security against attacks has grown. One of the many measures taken to this end was the adoption of various regulations that establish measures for the protection of critical infrastructures. All these regulations agree to include hydraulic infrastructures (both drinking water and waste water) in the catalog of critical infrastructure.

Among the many vulnerabilities that can be found in water distribution networks, telecontrol systems have the added risk of suffering not only physical failures but also cyber attacks. Therefore, any strategy that allows establishing a system of control of the quality of the received signals supposes an important contribution to the problem. This work tries to establish a strategy of control of the signals received by the telecontrol system so that before a fortuitous or intentional failure, it is detected and corrected in time.

In the case of the C-Town network, it is essential to obtain an analysis that allows to know the real state of the system, due to the fact that in the last years it was possible to implement a novel technology for the collection of data at a distance through the process that they carry out the respective field sensors and the actuator remote control. However, after this technology was introduced, anomalous tank levels began to be detected, for example, overflowing in some of the tanks unexpectedly while the water level readings were always below the alarm thresholds and The pumping maneuvers also appeared to be normal.

That is why this project seeks to provide protection to the C-TOWN supply network, thus creating an environment of certainty regarding the behavior of the system that depends largely on the measures taken by the corresponding apparatus, in this case it has An hourly register of some variables, which will be used to create a model that can detect inconsistencies in critical points of the network, whether accidental or provoked.

The network provides a set of pre-fault telemetry data, ie the above values without any manipulation, failure or accident in a period of 1 year. These data describe a normal behavior of the system and therefore are the ones that will be used to create the detection model. In addition, anomalous data are provided where values are incorrectly recorded for a

period of time equal to the previous one. Precisely the comparison of the behavior of these data sets will result in a methodology for the detection of anomalies in the data provided by telemeasurement, so that we will be able to determine true behavioral ranges of the hydraulic elements and also non-unusual behavior Of the same that warn about a possible failure.

To carry out this proposal will be necessary to build a mathematical model where it is possible to enter the data recorded and it can determine if they are correct or not, it will also be necessary to estimate the demands and therefore the patterns of consumption in each period of time, due to That the network lacks billing data therefore it is imperative to know what the true supply is at the respective time.

On the other hand, it will be necessary to make quality tests of the measurements obtained in the network to be sure that the information supplied is correct and that can be processed as a basis for the real understanding of the system. This leads us to propose not only tests of the true data, but also to evaluate the data set containing known or caused failures due to the creation of simulated fault environments using the EPANET program.

Finally, we will collect and evaluate the information obtained through a confusion matrix, which contains a table of double entry binary operation. The response of the detection system (safe situation or anomalous situation) with the effective operation of the telecontrol system will be checked on the one hand. If both answers match, there will be hits (positive if there is a fault and negative if there is not). In case there are discrepancies between the prediction of the detection system and the operation of the network, there will be a system failure (false positive if the system warns and there is no fault or false negative if there is a fault and it has not been warned). These discrepancies will serve to measure the reliability of the detection system. The columns of the array represent the instances of a real class while the rows represent the instances of a predicted class. This matrix will in turn result in two system states, FAIL or SAFE status, which will guarantee clarity about the behavior of the system when any type of anomaly described above occurs.

RESUM

En Valenciano: La comprensió, l'anàlisi i el millorament dels processos relacionats amb els abastiments urbans s'ha convertit en un dels principals objectius de la majoria de països a nivell mundial. Un dels principals interessos d'estos mateixos radica en el bon aprofitament del recurs que cada vegada és més escàs, així com la protecció de la infraestructura i un bon control del consum dels abonats. L'ocupació de la teledividida i el telecontrol de les xarxes facilita l'operació de les mateixes, donant respostes eficaces a canvis en les circumstàncies de funcionament.

Hui en dia i sobretot arran del ressorgiment de determinats grups terroristes de caràcter global ha crescut la preocupació per la seguretat enfront d'atacs. Una de les moltes mesures preses a este efecte va ser l'aprovació de diverses normatives en què s'establixen mesures per a la protecció de les infraestructures crítiques. Totes estes normatives coincidixen a incloure les infraestructures hidràuliques (tant d'aigua potable com d'aigües residuals) en el catàleg d'infraestructures crítiques.

D'entre els molts punts vulnerables que es poden trobar en les xarxes de distribució d'aigua, els sistemes de telecontrol tenen com a risc afegit la possibilitat de patir no sols fallades físiques, sinó també atacs cibernètics. Per tot això, qualsevol estratègia que permeta establir un sistema de control de la qualitat dels senyals rebudes suposa una important aportació al problema. Este treball pretén establir una estratègia de control dels senyals rebudes pel sistema de telecontrol de manera que davant d'una fallada fortuïta o intencionat, este siga detectat i corregit a temps.

En el cas de la xarxa C-Town, és imprescindible obtindre una anàlisi que permeta conèixer l'estat real del sistema, pel fet que en els últims anys es va aconseguir implementar una tecnologia nova per a l'arreglada de dades a distància a través del procés que duen a terme els sensors de camps respectius i el control remot d'actuadors. No obstant això, després que esta tecnologia fóra introduïda es va començar a detectar nivells anòmals en els tancs, per exemple, rebosamiento en alguns dels tancs de forma inesperada mentres que les lectures del nivell d'aigua sempre van estar per davall dels llindars d'alarma i les maniobres de bombament també pareixien ser normals.

És per això que este projecte busca brindar protecció a la xarxa d'abastiment C-TOWN, creant així un entorn de certesa quant al comportament del sistema que depén en gran part de les mesures preses pels aparells corresponents, en este cas es compta amb un registre hora a hora d'algunes variables, les quals s'usaren per a crear un model que puga detectar inconsistències en punts crítics de la xarxa, ja siguen de caràcter accidental o provocats.

La xarxa proporciona un conjunt de dades de teledividida previs a les avaries, és a dir, els valors abans mencionats sense cap tipus de manipulació, fallada o accident en un període

d'1 any. Estes dades descriuen un comportament normal del sistema i per tant són els que s'usaran per a la creació del model de detecció. A més es proporcionen dades anòmales on es registren d'una manera incorrecta valors durant un període de temps igual a l'anterior. Precisament la comparació del comportament d'estos conjunts de dades donarà com resultat una metodologia per a la detecció d'anomalies en les dades proporcionats per la tele mesura, per la qual cosa serem capaços de determinar rangs de comportament veridaders dels elements hidràulics i així mateixos comportaments no inusuals dels mateixos que alerten sobre una possible fallada.

Per a dur a terme esta proposta serà necessari construir un model matemàtic on siga possible ingressar les dades registrats i este puga determinar si són correctes o no, també serà necessari estimar les demandes i per tant els patrons de consums en cada període de temps, pel fet que la xarxa no té dades de facturació per tant és imprescindible saber quin és el subministrament verdader en l'instant respectiu. D'altra banda, serà necessari fer proves de qualitat dels mesuraments obtingudes en la xarxa per a estar segurs que la informació subministrada és correcta i que pot ser processada com a fonament de la comprensió real del sistema. Açò ens porta així mateix a plantejar no sols proves de les dades veridaderes, sinó que posteriorment a açò, avaluar el conjunt de dades que conté fallades coneguts o provocats pel fet que es plantejara entorns de fallades simulats, usant el programa EPANET.

Finalment, s'arreglarà i s'avaluarà la informació obtinguda a través d'una matriu de confusió, la qual conté una taula d'operació binària de doble entrada. Es contrastarà d'una banda la resposta del sistema de detecció (situació segura o situació anòmala) amb el funcionament efectiu del sistema de telecontrol. Si ambdós respostes coincidixen es tindran encerts (positiu si hi ha fallada i negatiu si no n'hi ha). En el cas que hi haja discrepàncies entre la predicció del sistema de detecció i el funcionament de la xarxa es tindrà una fallada del sistema (fals positiu si el sistema avisa i no hi ha fallada o fals negatiu si hi ha fallada i no ha sigut avisat). Estes discrepàncies serviran per a mesurar la fiabilitat del sistema de detecció. Les columnes de la matriu representen les instàncies d'una classe real mentres que les files representen les instàncies d'una classe predita. Esta matriu donarà com resultat al seu torn dos estats del sistema, estat de FALLADA O SEGUR, la qual cosa garantirà claredat sobre el comportament del sistema quan es presente qualsevol tipus d'anomalia descrites anteriorment.



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA

máster en ingeniería
hidráulica y medio ambiente
mihma



A Dios, porque siempre cumple sus promesas

A mis padres porque han sido el soporte de mi vida

A mi novia, por la calidez de su amor

A mis hermanos por su apoyo incondicional

A mi Tía, por bendecirme con esta oportunidad



AGRADECIMIENTOS

A mi director de Trabajo de Fin de Master, Dr. Javier Martínez Solano y codirector Dr. Pedro Iglesias Rey, mi más sincera gratitud por su confianza y apoyo para llevar a cabo este proyecto.

A mis compañeros del Master de Ingeniería Hidráulica y Medio Ambiente, por su amistad brindada.

A mis Padres, Gildardo y Neira por su respaldo incondicional durante este proceso de desarrollo profesional.

A mi Tía Ana del Rosario, por brindarme su cariño y brindarme esta oportunidad de cumplir este sueño.

A todos ellos, mi reconocimiento, respeto y gratitud



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA

máster en ingeniería
hidráulica y medio ambiente
mihma



INDICE

1. INTRODUCCION.....	20
1.1 JUSTIFICACION.....	20
1.2 ANTECEDENTES Y ALCANCE DEL TRABAJO.....	21
1.3 OBJETIVOS.....	22
1.3.1 OBJETIVO GENERAL.....	22
1.3.2 OBJETIVOS ESPECIFICOS.....	22
1.4 ESTRUCTURA DEL DOCUMENTO.....	22
2. INTRODUCCION AL PROBLEMA.....	25
2.1 DESCRIPCION DE LA RED COMO BENCHMARK.....	25
2.2 DESCRIPCION DEL PROBLEMA.....	27
2.3 PATRON DE CONSUMOS.....	27
2.4 CONTROL DE BOMBAS POR CONSIGNAS.....	28
2.5 CRITERIOS DE EVALUACION DEL ALGORITMO.....	28
2.6 DISPONIBILIDAD DE DATOS.....	29
3. CONSTRUCCION DEL MODELO.....	33
3.1 INTRODUCCIÓN.....	33
3.2 DETERMINACIÓN DE LAS DEMANDAS Y PATRONES DE CONSUMO.....	33
3.3 DETERMINACION DE LAS CURVAS CARACTERISTICAS DE LAS BOMBAS.....	45
3.4 DETERMINACION DE LOS CONTROLES.....	51
3.5 ANALISIS DE LOS TANQUES.....	52
4. COMPROBACION DE REGISTROS.....	59
4.1 VALIDACION DE CONSUMOS.....	59
4.2 VALIDACION DE LAS ESTACIONES DE BOMBEO.....	68
4.3 VALIDACION DE LOS TANQUES.....	76
5. RESULTADOS.....	81
5.1 ANOMALIAS EN LOS CONSUMOS.....	81
5.2 ANOMALIAS EN LAS BOMBAS.....	84
5.3 ANOMALIAS EN LOS TANQUES.....	88
5.4 MATRIZ DE RESULTADOS.....	90
6. MEJORAS DEL MODELO.....	93
6.1 ANALISIS DE LAS PRESIONES.....	93



6.2 ANALISIS DE CAUDAL BOMBEADO	110
6.3 CAUDAL DE LA VALVULA	117
6.4 RESULTADO DE LAS MEJORAS AL MODELO	119
7. CONCLUSIONES.	124
8. DESARROLLOS FUTUROS.....	125
9. BIBLIOGRAFÍA.....	126

LISTA DE FIGURAS

Figura 1. Modelo completo de la red C-Town en EPANET	25
Figura 2. Matriz de confusión para evaluación del Modelo.	29
Figura 3. Líneas de caudal para los Sectores de la red	34
Figura 4. Nivel del Tanque 3 asociado al paro de la bomba.	36
Figura 5. Nivel del Tanque 3 asociado al arranque de la bomba.	37
Figura 6. Curva de Modulación Sector 2	38
Figura 7. Nivel del Tanque 4 asociado a los arranques y paradas de la bomba.	38
Figura 8. Curva de Modulación Sector 3	39
Figura 9. Nivel del Tanque 5 asociado al arranque de la bomba.	40
Figura 10. Curva de Modulación Sector 4.....	40
Figura 11. Nivel del Tanque 7 asociado a los arranques y paradas de la bomba.	41
Figura 12. Curva de Modulación Sector 5.....	42
Figura 13. Flujo del caudal del Sector 1 a los demás Sectores	43
Figura 14. Nivel de los Tanques 1 y 2 asociado a los arranques de las bombas.	44
Figura 15. Curva de Modulación Sector 1.....	44
Figura 16. Estación de bombeo Sector 1.	45
Figura 17. Curva característica de la bomba 1	46
Figura 18. Curva característica de la bomba 2.....	47
Figura 19. Curva característica de la bomba 4.....	47
Figura 20. Curva característica de la bomba 6.....	48
Figura 21. Curva característica de la bomba 7	48
Figura 22. Curva característica de la bomba 8.....	49
Figura 23. Curva característica de la bomba 10	49
Figura 24. Curva característica de la bomba 11	50
Figura 25. Comportamiento de las bombas respecto al nivel de los tanques.....	51
Figura 26. Niveles del Tanque 1 dataset 1.....	53
Figura 27. Nivel del Tanque 2 dataset 1.....	53
Figura 28. Nivel del Tanque 3 dataset 1	54
Figura 29. Nivel del Tanque 4 dataset 1	55
Figura 30. Nivel del Tanque 5 dataset 1	56
Figura 31. Nivel del Tanque 6 dataset 1	57
Figura 32. Nivel del Tanque 7 dataset 1	57
Figura 33. Curva de modulación Sector 1 dataset 2.....	60
Figura 34. Coeficiente de Modulación Sector 1 dataset 2.	60
Figura 35. Coeficiente de Modulación normal Sector 1.	61
Figura 36. Curva de Modulación Sector 2 dataset 2.....	62
Figura 37. Coeficiente de modulación Sector 2 dataset 2.	62
Figura 38. Coeficiente de Modulación normal Sector 2.	63
Figura 39. Curva de Modulación Sector 3 dataset 2.....	64
Figura 40. Coeficiente de Modulación Sector 3 con anomalías.	64
Figura 41. Coeficiente de Modulación normal Sector 3.	65



Figura 42. Curva de Modulación del Sector 4 dataset 2.	65
Figura 43. Coeficiente de Modulación Sector 4 dataset 2.	66
Figura 44. Coeficiente de Modulación normal Sector 4	66
Figura 45. Curva de Modulación Sector 5 dataset 2.....	67
Figura 46. Coeficiente de Modulación Sector 5 dataset 2.	67
Figura 47. Coeficiente de Modulación normal Sector 5	68
Figura 48. Curva característica de la bomba 1 con anomalías.	69
Figura 49. Curva característica de la bomba 2 con anomalías.	69
Figura 50. Curva característica de la bomba 4 dataset 2.	70
Figura 51. Curva característica bomba 6 con anomalías.	71
Figura 52. Curva característica de la bomba 7 con anomalías.	72
Figura 53. Variación de la velocidad de giro de la bomba 7.	73
Figura 54. Curva característica de la bomba 8 con anomalías.	74
Figura 55. Curva característica de la bomba 10 con anomalías.	75
Figura 56. Curva Característica de la bomba 11 con anomalías.....	75
Figura 57. Nivel del Tanque 1 dataset 2	76
Figura 58. Nivel del tanque 2 dataset 2	77
Figura 59. Nivel del Tanque 3 dataset 2	77
Figura 60. Nivel del Tanque 4 dataset 2	78
Figura 61. Nivel del Tanque 5 dataset 2	78
Figura 62. Nivel del Tanque 6 dataset 2	79
Figura 63. Nivel del Tanque 7 dataset 2	79
Figura 64. Localización de Aciertos Patrón 1.	82
Figura 65. Localización de Aciertos Patrón 3	83
Figura 66. Localización de Aciertos Patrón 5	83
Figura 67. Localización de aciertos bomba 1	85
Figura 68. Localización de Aciertos bomba 2.....	85
Figura 69. Localización de Aciertos bomba 6.....	86
Figura 70. Localización de Aciertos bomba 6.....	86
Figura 71. Localización de Aciertos bomba 8.....	87
Figura 72. Localización Aciertos bomba 10	87
Figura 73. Localización Aciertos bomba 11	88
Figura 74. Localización Aciertos Tanque 1.....	89
Figura 75. Presiones a la entrada de la Estación de bombeo 1 dataset 1.	94
Figura 76. Presiones a la salida de la Estación de bombeo 1 dataset 1.	95
Figura 77. Presión a la entrada de la Estación de bombeo 1 dataset 2.	95
Figura 78. Presión a la salida de la Estación de bombeo 1 dataset 2.	96
Figura 79. Estación de bombeo Sector 2	96
Figura 80. Presión a la entrada de la estación de bombeo 2 dataset 1	97
Figura 81. Presión a la salida de la Estación de bombeo 2 dataset 1	97
Figura 82. Presión a la entrada de la Estación de bombeo 2 dataset 2	98
Figura 83. Presión a la salida de la Estación de bombeo 2 dataset 2	98
Figura 84. Estación de bombeo Sector 3.	99
Figura 85. Presión a la entrada de la Estación de bombeo 3 dataset 1	99
Figura 86. Presión a la salida de la Estación de bombeo 3 dataset 1.	100
Figura 87. Presión a la entrada de la Estación de bombeo 3 dataset 2.	100



Figura 88. Presión a la salida de la Estación de bombeo 3 dataset 2	101
Figura 89. Estación de bombeo Sector 4	101
Figura 90. Presión a la entrada de la Estación de bombeo 4 dataset 1	102
Figura 91. Presión a la salida de la Estación de bombeo 4 dataset 1.	102
Figura 92. Presión a la entrada de la Estación de bombeo 4 dataset 2	103
Figura 93. Presión a la salida de la Estación de bombeo 4 dataset 2	103
Figura 94. Estación de bombeo 5.....	104
Figura 95. Presión a la entrada de la Estación de bombeo 5 dataset 1	105
Figura 96. Presión a la salida de la Estación de bombeo 5 dataset 1.	105
Figura 97. Presión a la entrada de la estación de bombeo 5 dataset 2	106
Figura 98. Presión a la salida de la Estación de bombeo 5 dataset 2	106
Figura 99. Válvula de Regulación.....	107
Figura 100. Presión a la entrada de la Válvula dataset 1.....	108
Figura 101. Presión a la salida de la Válvula dataset 1.....	108
Figura 102. Presión a la entrada de la Válvula dataset 2.....	109
Figura 103. Presión a la salida de la Válvula dataset 2.....	109
Figura 104. Caudal de la bomba 1 dataset 1.....	111
Figura 105. Caudal de la bomba 2 dataset 1.....	111
Figura 106. Caudal de la bomba 1 dataset 2.....	112
Figura 107. Caudal de la bomba 2 dataset 2.....	112
Figura 108. Caudal de la bomba 4 dataset 1.....	113
Figura 109. Caudal de la bomba 4 dataset 2.....	113
Figura 110. Caudal de la bomba 7 dataset 1.....	114
Figura 111. Caudal de la bomba 7 dataset 2.....	114
Figura 112. Caudal de la bomba 8 dataset 1.....	115
Figura 113. Caudal de la bomba 8 dataset 2.....	116
Figura 114. Caudal de la bomba 10 dataset 1.....	116
Figura 115. Caudal de la bomba 10 dataset 2.....	117
Figura 116. Caudal de la Válvula dataset 1.....	118
Figura 117. Caudal de la Válvula dataset 2.....	118
Figura 118. Diagrama de efectividad de las sospechas.	123



LISTA DE TABLAS

Tabla 1. Distribución de elementos hidráulicos en cada Sector.....	26
Tabla 2. Descripción de los periodos de anomalías.....	30
Tabla 3. Formato de Registros telemando para los Tanques.....	30
Tabla 4. Formato de registro telemando para las estaciones de bombeo.....	31
Tabla 5. Formato de registro telemando para las presiones.....	31
Tabla 6. Formato de registro telemando para fallos.....	32
Tabla 7. Ecuaciones de las curvas características en todas las bombas.....	50
Tabla 8. Propiedades de los Tanques Sector 1.....	52
Tabla 9. Propiedades del Tanque Sector 2.....	54
Tabla 10. Propiedades del Tanque 4 Sector 3.....	54
Tabla 11. Propiedades del Tanque 5 Sector 4.....	55
Tabla 12. Propiedades del Tanque 6 y 7 dataset 1.....	56
Tabla 13. Rango de funcionamiento de las bombas Sector1.....	69
Tabla 14. Rango de funcionamiento de las bombas Sector 2.....	70
Tabla 15. Rango de funcionamiento de las bombas Sector 2.....	71
Tabla 16. Rango de funcionamiento de las bombas Sector 2.....	73
Tabla 17. Rango de funcionamiento de las bombas Sector 2.....	74
Tabla 18. Resultados de la detección por Consumos.....	81
Tabla 19. Resultados de la detección por bombeo.....	84
Tabla 20. Resultado de la detección en los tanques.....	89
Tabla 21. Números generales de detección.....	90
Tabla 22. Matriz de resultados del Modelo de detección.....	90
Tabla 23. Resultados de la detección por Presiones.....	119
Tabla 24. Resultados finales de la detección por bombeo.....	119
Tabla 25. Mejorías en las estaciones de bombeo.....	120
Tabla 26. Resultados de la detección por Caudal.....	120
Tabla 27. Resultados finales de la detección por Patrones de consumo.....	121
Tabla 28. Mejoría en los Patrones de Consumo.....	121
Tabla 29. Números finales de detección.....	122
Tabla 30. Matriz final de resultados del Modelo de detección.....	122



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA

máster en ingeniería
hidráulica y medio ambiente
mihma

1. INTRODUCCION

España y el mundo entero hoy en día han sido testigos del impacto de los ciberataques, que con la aparición en estas últimas semanas del famoso WannaCry ha disparado las alarmas en todas las organizaciones gubernamentales y privadas a nivel mundial. Este ciberataque a secuestrado la información de centenares de ordenadores pidiendo rescates de más de 300 dólares por cada uno. El instituto Nacional de Ciberseguridad (Incibe) sitúa a España como el tercer país más atacado en el mundo detrás de Estados Unidos y Reino Unido. (MUNDO, 2016)

Estos ciberataques se basan principalmente en buscar una forma de entrar a una plataforma privada a través de correos electrónicos y aplicaciones aparentemente no dañinas, que terminan después de instalarse, bloqueando el acceso a la información del sistema. En el caso de WannaCry, después de apropiarse del sistema, podía modificar o eliminar completamente la información de modo que muchas organizaciones vieron viable la posibilidad de pagar el “rescate” de su información a través del pago sugerido.

Particularmente en España se han confirmado más de 1.200 ataques, donde uno de los más evidentes fue a la sede telefónica en Madrid, donde los trabajadores han optado por apagar los ordenadores para evitar daños mayores. También Wanna Cry infectó desde los equipos de 16 hospitales y centros de salud en Reino Unido, hasta los de la empresa de automóviles Renault, en Francia, que se ha visto obligada a parar la producción de varias plantas en el país, pasando por los del Ministerio del Interior ruso.

El reciente ciberataque a gran escala "sirve para enviar un mensaje muy claro: todos los sectores son vulnerables y deben tomarse absolutamente en serio la necesidad de funcionar con sistemas actualizados de protección" afirmó el director de Europol. Es por eso que las organizaciones a nivel mundial concuerdan en aumentar recursos destinados a la protección de los ataques cibernéticos. (País, 2017)

1.1 JUSTIFICACION

Las redes de abastecimiento de hoy en día requieren cada vez más sistemas de control efectivos que puedan garantizar un buen nivel de protección y comprensión acerca de todos los procesos llevados a cabo durante su operación. Es por eso que, con el aumento de grupos terroristas a nivel mundial, capaces de llevar a cabo operaciones en contra del bienestar social, se ha priorizado la

elaboración de metodologías para la detección y prevención de anomalías, en este contexto, la elaboración de modelos y estrategias capaces de detectar fallos cibernéticos. (Oliker, 2015)

Con el surgimiento de estas graves anomalías en las redes de distribución de agua en los últimos años, la información acerca de cómo hacer frente a estos tipos de irrupciones es muy poca y no solo eso, sino que no existe datos concretos sobre los daños que puede causar a corto y largo plazo, de esta forma, las organizaciones a nivel mundial han intentado mejorar cada año las estrategias de operación y vigilancia de las redes (Institute, 2017). En cuanto a este trabajo, los resultados ayudaran a entender mejor el comportamiento de cualquier red que está sometida a manipulaciones externas, de forma que sea posible caracterizar estas anomalías y abrir un enfoque de desarrollo posterior para el mejoramiento de las condiciones de la red.

En este sentido, la conservación del Recurso que cada vez es más escaso en las áreas metropolitanas adquiere un valor muy importante, no solo a nivel de la satisfacción de las demandas, sino desde el punto de vista ambiental, que entre otras cosas ha llevado a la firma de acuerdos internacionales para el buen uso del agua (McCaffrey, 2009). De igual forma, este trabajo busca evitar el desaprovechamiento del recurso causado por el desbordamiento de Tanques de almacenamiento, así como la protección de los elementos hidráulicos y la seguridad del sistema.

1.2 ANTECEDENTES Y ALCANCE DEL TRABAJO

El “Sistema Ciber – Físico” o como se le denomina en inglés “Cyber Physical System” (CPS) es el término usado por la fundación americana NSF (National Science Foundation) que toma nombre del resultado de dotar a los componentes u objetos físicos que nos encontramos de forma habitual en nuestro entorno de trabajo, de capacidades de computación y de comunicación para convertirlos en objetos inteligentes. Estos permiten así, superar a los simples sistemas empotrados actuales en cuanto a capacidad, seguridad, escalabilidad, adaptabilidad, resiliencia y usabilidad, pudiendo trabajar en conjunto formando ecosistemas distribuidos y totalmente autónomos. (Mundo, 2015)

Los sistemas Ciber-Físicos, son parte del mundo moderno y podemos observar estos procesos a través de los semáforos y la red eléctrica, donde los elementos físicos e integrados trabajan en conjunto para realizar acciones de una forma más eficiente. De la misma manera, en las redes de distribución de agua cada vez los procesos que tienen que ver con bombas, válvulas, tanques y embalses son más automatizados, provocando así un interés por parte de varias organizaciones como es el caso de Environmental & Water Resources Institute y University of Exeter, las cuales entre otras han venido desarrollando modelos enfocados no solamente a la detección de anomalías en redes de distribución de agua, sino que también han participado en desafíos desde hace más de dos décadas, por ejemplo “The Battle of the Water Networks (BWN)” llevada a cabo en 1985 , luego “Battle of the Water Sensor Networks (BWSN)” en el 2006 y “Battle of the Water Calibration Networks (BWCN)” en el 2010. (Institute, 2017)



En este año Battle of the Attack Detection Algorithms (BATADAL) ha enfocado su trabajo en la detección de anomalías a través de modelos matemáticos para la red C-Town, generando así una gran expectativa en muchas organizaciones que hoy en día dependen de sistemas integrados para la lectura y comprensión del estado de cada una de sus redes de abastecimiento.

En el caso del presente trabajo, su enfoque estará en la elaboración de un modelo de detección capaz de detectar anomalías en la red C-Town y pueda establecer, a partir de varios criterios, el estado del sistema para cada instante de tiempo proporcionado por el conjunto de datos de la red.

1.3 OBJETIVOS

1.3.1 OBJETIVO GENERAL

El objetivo general de este trabajo es la construcción de un modelo matemático capaz de detectar anomalías en la red C-Town a partir del análisis de los datos registrados por el sistema de telemando del abastecimiento y anticipar así anomalías en el mismo, ya sean accidentales o provocadas.

1.3.2 OBJETIVOS ESPECIFICOS

Para llevar a cabo este objetivo general se plantean los siguientes objetivos específicos:

Conocer el funcionamiento de la red, lo cual incluye describir el funcionamiento de las demandas, patrones de consumo y controles de las bombas correspondientes a cada sector de la red.

Presentar un planteamiento o metodología detallada del modelo de detección usado para la detección de anomalías.

Elaboración de criterios de análisis de detección de anomalías: los patrones de consumo, las curvas características de las bombas, los niveles de los tanques y otras variables que se requieran.

1.4 ESTRUCTURA DEL DOCUMENTO

Este trabajo comenzará en el capítulo 2 con una descripción completa de la red de estudio, es decir, la caracterización del funcionamiento y los elementos hidráulicos involucrados en el proceso de abastecimiento de C-Town. También se explicará el uso de esta red para el desarrollo de nuevas estrategias de detección de anomalías por parte de organizaciones internacionales. A partir de esto



se planteará el problema a solucionar, involucrando el contexto real del sistema y la información que se cuenta y se desconoce para la elaboración del modelo de detección. Con esto se explicará los enfoques del modelo partiendo de las demandas y las estaciones de bombeos para de esta forma establecer los criterios de evaluación finales que determinaran el estado general del Sistema.

A continuación, en el capítulo 3 se comenzará con la elaboración del Modelo a partir de los enfoques previos, analizando el primer conjunto de datos proporcionado y colocando de esta manera la base de comparación para la detección de fallos. En este punto se determinarán los consumos de la red, el funcionamiento correcto de las estaciones de bombeo y los niveles adecuados en cada Sector.

Terminado esto, en el capítulo 4 se realizará una comprobación del sistema cuando está sometido a ataques con el uso del segundo conjunto de datos y de esta forma detectar comportamientos anormales en cuanto al consumo de los usuarios, funcionamiento de las estaciones de bombeo y dinámica de los tanques de la red.

En el capítulo 5 se presentará los resultados del análisis del sistema a partir del segundo conjunto de datos, caracterizando las anomalías y evaluando la eficiencia de cada criterio respecto al objetivo del Modelo. Se clasificará por lo tanto las detecciones a partir de la matriz de confusión y se obtendrá un concepto global del estado del sistema a partir de los resultados encontrados.

Por último en el capítulo 6, se propondrán mejoras para el modelo, de tal manera que se pueda conseguir los mejores resultados para la red, atendiendo a la elaboración de nuevos criterios y nuevos enfoques en el sistema. En consecuencia, se conocerán los resultados finales a partir de estas mejoras y se dará a conocer las conclusiones del análisis, proyectando el trabajo hacia futuros estudios que lleven a un mejor desarrollo del tema.



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA

máster en ingeniería
hidráulica y medio ambiente
mihma

2. INTRODUCCION AL PROBLEMA

2.1 DESCRIPCION DE LA RED COMO BENCHMARK

C-Town contiene 388 conexiones de los cuales 334 son puntos de consumo y están repartidos en 5 sectores; es una red de tamaño mediano real con dos tipos de almacenamientos principales (T1 y T2), y una estación de bombeo principal (S1) que consta de tres bombas (PU1, PU2 y PU3). Las operaciones de PU1 y PU2 están reguladas por los niveles de agua en T1, mientras que PU3 se mantiene fuera de operación. Los cinco tanques restantes son rellenos por cuatro estaciones secundarias de bombeo que inyectan el agua desde T1 y T2. Mientras T1 está conectado directamente a S1, la rama correspondiente a T2 está conectada a la fuente a través de una válvula de regulación (V2) controlado por el nivel de agua en T2. La figura de la red se muestra a continuación:

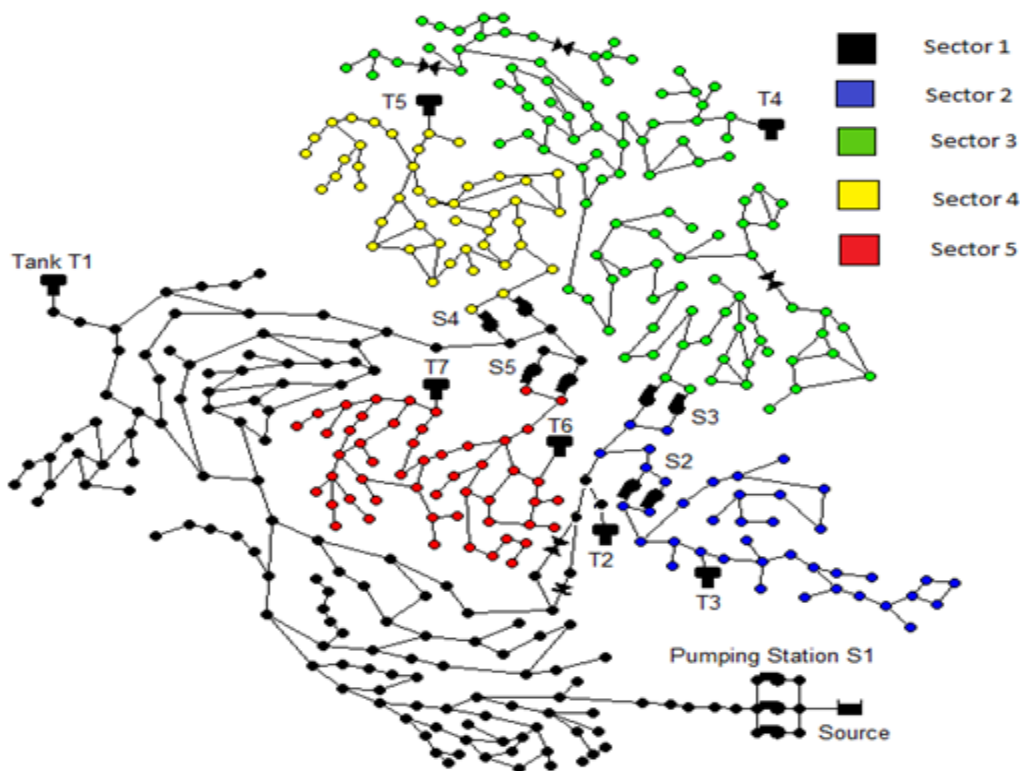


Figura 1. Modelo completo de la red C-Town en EPANET

Cada sector posee uno o dos tanques que controlan cada estación de bombeo, por lo que primero habrá que identificar cuáles son los tanques asociados a cada estación y de esta manera separar cada zona con sus respectivos elementos hidráulicos. A continuación, se muestra la tabla de cada Sector con sus correspondientes elementos:

SECTOR	TANQUE	BOMBAS
1	T1,T2	PU1,PU2,PU3
2	T3	PU4,PU5
3	T4	PU6,PU7
4	T5	PU8,PU9
5	T6,T7	PU10,PU11

Tabla 1. Distribución de elementos hidráulicos en cada Sector.

Ahora, con la implementación de la nueva tecnología de captación de datos, se ha podido obtener medidas hora a hora de los siguientes elementos:

- Nivel de agua en cada tanque
- Estado (0 para OFF / CLOSED, 1 para ON / OPEN) para cada bomba y válvula en el sistema
- Flujo a través de cada bomba y válvula
- Presión de succión y descarga para cada válvula y estación de bombeo

Estas variables se encuentran disponibles en dos conjuntos de datos, es decir, registros hora a hora del comportamiento de los elementos hidráulicos. El primer conjunto de datos contiene registros correctos de un año, mientras que el segundo conjunto de datos cuenta con mediciones anómalas tomadas desde el 04 de Julio hasta el 25 de diciembre del mismo año. De igual forma se cuenta con un archivo en EPANET de la red, la cual contiene la topología del terreno, la geometría de los elementos hidráulicos y la forma original del sistema. Tanto los dos conjuntos de datos como el archivo suministrado, serán la base para la elaboración del algoritmo de detección de fallos.

También es importante mencionar que esta red ha servido los últimos años como referencia para varios proyectos de diseño y operación hidráulica, donde a partir de una red existente se planteó la posibilidad de actualizaciones para demandas futuras y nuevas áreas de desarrollo. El diseño llevo a la adición de tuberías nuevas y paralelas, almacenamientos, controles para bombas y válvulas y la implementación de una nueva fuente de alimentación. Luego, con la adquisición de nuevas tecnologías para el mejoramiento de la red, el desafío se ha ido generando de la misma manera, buscando siempre buenos algoritmos que produzcan resultados eficientes, de tal forma que hoy en día la última implementación ha sido la utilización de sistemas de tele medida para la monitorización de la red. (Marchi, 2014).

2.2 DESCRIPCION DEL PROBLEMA

BATADAL o batalla de algoritmos de detección de Ataques, es una iniciativa respaldada por la ASCE (American Society of civil Engineers) y propuesta para evaluar el rendimiento de los algoritmos para la detección de anomalías cibernéticas en los sistemas de distribución de agua. Esta organización desafía a los participantes a aportar un algoritmo de detección de anomalías para la red C-Town, siguiendo un conjunto de reglas que determinan el objetivo exacto del proyecto. (Institute, 2017).

El fundamento es que los sistemas de distribución de agua modernos se basan en los ordenadores, sensores y actuadores, tanto para fines operativos como de vigilancia. Esta combinación de procesos físicos y los sistemas integrados (sistemas ciber-físicos) mejora el nivel de servicio de las redes de distribución de agua, pero los expone a las amenazas potenciales de ataques cibernéticos. Durante los últimos años, varios sistemas de abastecimiento y distribución de agua han sido atacados, llevando a si a la creación de agencias de seguridad cibernética y asociaciones internacionales para defender las redes de agua. Sin embargo, existe poca información sobre el efecto potencial de estas anomalías, así como del diseño e implementación de algoritmos de detección que identifican comportamientos anómalos de sensores, bombas y otros componentes de las redes de agua.

Servicios Públicos C-Town es el principal operador del sistema de distribución de agua de C-Town. Durante muchos años, esta entidad ha operado una topología de distribución estática. En el último año, se ha introducido nueva tecnología inteligente para permitir la recopilación de datos a distancia de los sensores en el campo, y el control remoto de actuadores. Tiempo después de que la nueva tecnología se introdujo, se observaron bajos niveles anómalos en el tanque T5 y altos niveles en el tanque T1. Un mes después, ocurrió un desbordamiento de agua en el tanque T1. Mientras que el personal en el centro de control fue capaz de ver las lecturas anómalas durante los dos primeros episodios, el tanque T1 desbordó de forma inesperada mientras que las lecturas del nivel del agua eran siempre por debajo de los umbrales de alarma y las operaciones de bombeo parecían ser normales. Esto llevo a levantar sospechas sobre posibles anomalías cibernéticas para todos estos episodios; en conclusión, se empezó a considerar agentes exteriores capaces de activar y desactivar los actuadores en C-Town, así como la alteración de las lecturas de los sensores desplegados en la red y el estado de informe de actuadores, interfiriendo así con las conexiones establecidas entre los componentes conectados en red (Nils Ole Tippenhauer, 2016).

2.3 PATRON DE CONSUMOS

Debido a que el sistema sólo cuenta con los datos mencionados anteriormente, es necesario por lo tanto entender y cuantificar el comportamiento de la red respecto al consumo de los usuarios,

puesto que es una red de abastecimiento que no cuenta con contadores que determinen el caudal que la ciudad consume durante el tiempo establecido en los conjuntos de datos proporcionados. Esto muy importante a la hora de comprender el funcionamiento de la red, puesto que existirán momentos donde el consumo de los abonados será muy bajo y por ende el caudal bombeado será menor, de forma que se creará un criterio de evaluación a partir de la elaboración de estos patrones de consumo.

De esta manera se deberá elaborar patrones de consumo en cada sector que puedan aportar la información necesaria para separar el caudal que realmente está siendo consumido, con aquel que puede generarse a consecuencia de alguna anomalía. El resultado de estos será representado a través de curvas de modulación.

2.4 CONTROL DE BOMBAS POR CONSIGNAS.

La red contiene en total 5 sectores como ya se ha mencionado anteriormente y además el funcionamiento de cada estación de bombeo está controlado indirectamente por el estado del Tanque asociado a dicha estación, por lo cual se analizará si esa relación puede brindar información adicional que pueda llevar a una posible detección de anomalías; esto a partir de comparaciones de vaciado y llenado de los tanques con estados de paro y encendido de las bombas, de tal manera que pueda llegar ser o no una alternativa viable de detección.

Por otro lado, también se necesario analizar el comportamiento individual de las bombas en las estaciones, de tal manera que se pueda describir y entender el funcionamiento a partir de sus curvas características respectivas, las cuales se construirán en base a los datos brindados de las presiones a la entrada y salida de cada una de ellas, de tal forma que la comparación de alturas y caudales sea un criterio importante para la detección de fallos.

Por último, se crearán simulaciones comparativas para los dos conjuntos de datos, de manera que tanto los caudales, alturas y presiones sean la base para elaboración del criterio en cuanto a las estaciones de bombeo.

2.5 CRITERIOS DE EVALUACION DEL ALGORITMO

BATADAL creó esta propuesta con la visión de encontrar el mejor algoritmo de detección de anomalías para C-Town, por lo cual adopta criterios cualitativos y cuantitativos basados en la clasificación de eventos o matriz de confusión (Powers, 2011).

- La matriz de confusión: Es una tabla utilizada para describir el rendimiento del algoritmo. Las columnas de la matriz representan las instancias de un estado real mientras que las filas muestran las instancias de un estado predicho. En BATADAL se emplea la matriz de confusión

para evaluar el desempeño de los algoritmos utilizando dos clases, *fallo* y *seguro*, que producen la matriz 2x2. Las celdas de la matriz de confusión se muestran en la Figura 2:

MATRIZ DE RESULTADOS		Estado Actual	
		FALLO (positivo)	SEGURO (Negativo)
Estado Predicho	FALLO (positivo)	VP	FP
	SEGURO (negativo)	FN	VN

Figura 2. Matriz de confusión para evaluación del Modelo.

Donde:

Verdadero Positivo (VP): el sistema está bajo ataque y el algoritmo lo reconoce.

Verdadero Negativo (VN): el sistema no está bajo ataque y el algoritmo lo reconoce.

Falso Negativo (FN): el sistema está siendo atacado pero el algoritmo no lo detecta.

Falso Positivo (FP): el sistema no está bajo ataque, pero el algoritmo detecta una amenaza inexistente (falsa alarma).

- Localización: El mecanismo de detección pueden localizar qué área de la red ha sido atacada e identificar los componentes objetivo. Esta información es valiosa para preparar una respuesta adecuada para resolver.

Teniendo en claro los criterios generales de la evaluación del modelo, es necesario mencionar además que el propósito es detectar si es posible todas las anomalías presentadas, es decir, maximizar el valor de VP, de modo que los falsos negativos o FN sean cero. En este sentido, se intentará que los valores adicionales como VN y FP sean los más óptimos posibles, pero con mayor prioridad los valores de VP y FN.

2.6 DISPONIBILIDAD DE DATOS.

Uno de los factores críticos en toda la descripción del problema está en la disponibilidad de los datos, debido a que red cuenta con escasa información proveniente de los autómatas y recolectados a través de la tele medida. En este punto, C- Town proporciona dos conjuntos de datos, que durante todo el desarrollo del proyecto llamaremos Dataset 1 y 2 respectivamente. El primero contiene datos

de 8761 horas tomadas desde el 6 de enero de 2014 hasta el 6 de enero de 2015, en total 1 año de registros que contienen el estado de cada uno de los elementos hidráulicos controlados por los autómatas y que presentan un comportamiento normal y habitual de la red. En el mismo sentido, el dataset 2 contiene 4177 horas de registros, tomadas desde el 4 de Julio de 2016 hasta el 25 de diciembre de 2016; este dataset tiene en total 5 periodos de fallos distribuidos durante toda la simulación cuya duración y hora de inicio se presentan a continuación:

PERIODO FALLO	HORA INICIO	HORA FINAL	DURACION (h)
1	14/09/16 03	15/09/16 20	42
2	09/10/16 09	11/10/16 20	60
3	30/10/16 19	01/11/16 07	37
4	27/11/16 04	27/11/16 10	7
5	06/12/16 21	09/12/16 21	73

Tabla 2. Descripción de los periodos de anomalías.

Para el dataset 1 y 2, los valores concernientes a los tanques están determinados por los niveles en cada instante, por lo cual como se muestra en la **Tabla 3**, los datos contabilizan la fecha del registro y los diferentes valores hora a hora del comportamiento de los tanques. A continuación, se presentan los registros de las primeras 12 horas para los niveles correspondientes.

DATETIME	L_T1	L_T2	L_T3	L_T4	L_T5	L_T6	L_T7
04/07/16 00	2.44	5.24	3.19	4.1	2.86	5.5	4.39
04/07/16 01	2.66	4.53	3.2	4.18	3.29	5.44	4.53
04/07/16 02	3.11	3.66	3.66	4.21	3.87	5.15	3.22
04/07/16 03	3.62	3.04	4.17	4.04	3.56	4.98	2.4
04/07/16 04	4.08	2.68	4.73	3.2	3.11	5.39	3.46
04/07/16 05	4.53	2.1	5.26	3.29	2.76	5.5	4.67
04/07/16 06	4.84	1.57	5.12	3.9	2.29	5.3	3.9
04/07/16 07	5.19	1.07	4.85	4.45	1.87	5	2.66
04/07/16 08	5.22	0.94	4.52	3.51	1.55	5.09	2.48
04/07/16 09	5.11	0.68	4.11	2.41	1.77	5.27	2.57
04/07/16 10	4.73	0.62	3.59	2.66	2.13	5.43	2.63
04/07/16 11	4.24	1.01	3.22	2.52	2.5	5.5	2.53
04/07/16 12	3.74	1.23	3.17	2.67	3.03	5.5	2.7

Tabla 3. Formato de Registros telemando para los Tanques

Del mismo modo, los dos dataset cuentan con los registros del caudal de las bombas, donde la información del telemando recoge la información del caudal bombeado, así como del estado de bomba, es decir, si está encendida y apaga. La **Tabla 4** muestra como ejemplo los registros para la estación de bombeo 1 para las primeras 12 horas.

DATETIME	F_PU1	S_PU1	F_PU2	S_PU2	F_PU3	S_PU3
04/07/16 00	93.63	1	93.65	1	0	0
04/07/16 01	89.41	1	89.43	1	0	0
04/07/16 02	89.88	1	89.89	1	0	0
04/07/16 03	88.1	1	88.12	1	0	0
04/07/16 04	87.01	1	87.03	1	0	0
04/07/16 05	110.09	1	0	0	0	0
04/07/16 06	107.71	1	0	0	0	0
04/07/16 07	110.23	1	0	0	0	0
04/07/16 08	111.38	1	0	0	0	0
04/07/16 09	112.78	1	0	0	0	0
04/07/16 10	117.5	1	0	0	0	0
04/07/16 11	118.65	1	0	0	0	0
04/07/16 12	118.91	1	0	0	0	0

Tabla 4. Formato de registro telemando para las estaciones de bombeo.

Por el lado de las presiones, en la **Tabla 5** se muestran los valores hora a hora de los puntos correspondientes a la entrada y salida de las estaciones de bombeo y válvula; por lo cual el centro de control tiene información de las caídas de presión que se generan durante los bombeos. A continuación, se muestran los datos de presión para las primeras 6 horas de registro.

DATETIME	P_J280	P_J269	P_J300	P_J256	P_J289	P_J415	P_J302	P_J306	P_J307	P_J317	P_J14	P_J422
04/07/16 00	2.98	36.38	28.26	70	28.22	85.87	21.69	82.72	21.58	71.99	39.33	29.64
04/07/16 01	2.98	38.6	24.39	87.73	24.45	84.87	29.81	86.62	29.81	59.76	42.17	26.15
04/07/16 02	2.98	38.36	23.84	89.29	23.9	87.11	29.85	87.64	29.85	58.5	42	25.56
04/07/16 03	2.98	39.29	27	91.98	27.1	68.75	31.6	64.25	31.47	72.3	43.24	28.38
04/07/16 04	2.98	39.85	26.67	92.11	26.76	68.74	32.3	64.23	32.17	72.53	44	28.04
04/07/16 05	2.99	27.41	22.59	91.94	22.64	86.26	30.41	63.36	30.3	79.47	39.76	24.28
04/07/16 06	2.99	28.74	24.78	75.29	24.74	87.03	33.98	63.13	33.98	59.47	41.24	26.13

Tabla 5. Formato de registro telemando para las presiones

Por último, en el caso del dataset 2, se cuenta con los fallos producidos durante todo el periodo de registro y como se ve en la **Tabla 6**, se tiene por un lado la fecha de la medida y por el otro el estado del sistema, es decir, si el valor es -999 indica que en ese instante el sistema está libre de fallos, mientras que si se tiene el valor de 1, entonces la red está en estado de fallo. La figura siguiente muestra el formato de anomalías de la red.

DATE TIME	ATT_FLAG
14/09/16 00	-999
14/09/16 01	-999
14/09/16 02	-999
14/09/16 03	1
14/09/16 04	1
14/09/16 05	1
14/09/16 06	1
14/09/16 07	1
14/09/16 08	1
14/09/16 09	1
14/09/16 10	1
14/09/16 11	1
14/09/16 12	1

Tabla 6. Formato de registro telemando para fallos

En conclusión, esta es la información suministrada por BATADAL para la elaboración del modelo, de modo que los valores han sido exportados directamente del telemando a una hoja de cálculo para su posterior análisis.

3. CONSTRUCCION DEL MODELO

3.1 INTRODUCCIÓN

En este punto del trabajo se abordará la elaboración del modelo matemático de detección de fallos para la red C-Town a partir de un planteamiento que dé solución a las problemáticas establecidas por BATADAL y aquellas que se puedan generar durante el proceso, generando así una mejor comprensión de los fallos del sistema. Para llevar a cabo esto, se usará como base de elaboración la información del dataset 1, puesto que como se había mencionado anteriormente, contiene información correcta de los elementos hidráulicos y en consecuencia en ausencia de fallos accidentales y provocados.

Para desarrollar este modelo primeramente se pensó en el gran inconveniente de los datos de consumo, puesto que no se tiene información acerca del comportamiento de los usuarios de C-Town. Para ello, primeramente será necesario calcular los patrones de consumo a partir de los datos que se cuenta, en este caso del dataset 1, los cuales no contienen fallos. De esta manera se tendrá 5 patrones de consumo asociados a cada sector de la red, generando por ultimo una base de comparación para la detección de fallos. Por otro lado, habrá que determinar los controles correctos de funcionamiento de la red, los cuales estarán asociados directamente al comportamiento de las bombas. Este análisis producirá consignas correctas que indiquen niveles normales de los tanques asociados a cada estación de bombeo.

Por último, se determinarán las curvas de cada una de las bombas, a partir de las presiones medidas a la entrada y salida de las estaciones de bombeo. El dataset 1 producirá curvas con características normales, que se compararan en su momento con las del dataset 2 que contienen fallos para que de esta manera se produzcan rangos de alturas y caudales correctos que sirvan para detectar anomalías en el sistema. Con todo esto se logrará crear una tabla donde aparecerán hora a hora el estado del sistema atendiendo a los criterios antes mencionados.

3.2 DETERMINACIÓN DE LAS DEMANDAS Y PATRONES DE CONSUMO.

Para poder determinar las demandas correspondientes, será necesario analizar cada sector por separado, debido a que los elementos hidráulicos varían de acuerdo al sector donde nos encontremos. Esto se realiza a partir de un balance de caudal, el cual se plantea con base al concepto de que todo el caudal que se pueda estar bombeando puede servir para llenar los tanques o para el suministro de los usuarios. Sin embargo, hay que considerar como un punto crítico los arranques y

paradas de las bombas, puesto que no se conoce el momento exacto cuando las bombas empiezan o detienen el bombeo, tan solo se cuenta el estado de hora en hora y esto llevara a plantear métodos alternativos de cálculo.

Por otro lado, como se observa en la **Figura 3**, el bombeo se realiza desde la estación de bombeo 1 y se reparte a los otro cuatro sectores de la red; en consecuencia, para poder determinar los consumos del Sector 1 se ha de calcular primeramente este valor en los demás sectores y con la suma de todos ellos, evidentemente apartar el consumo real del Sector en cuestión. Eso significa que el cálculo de las demandas y el patrón de consumo se iniciaran desde el Sector 2 y terminaran con los del Sector 1.

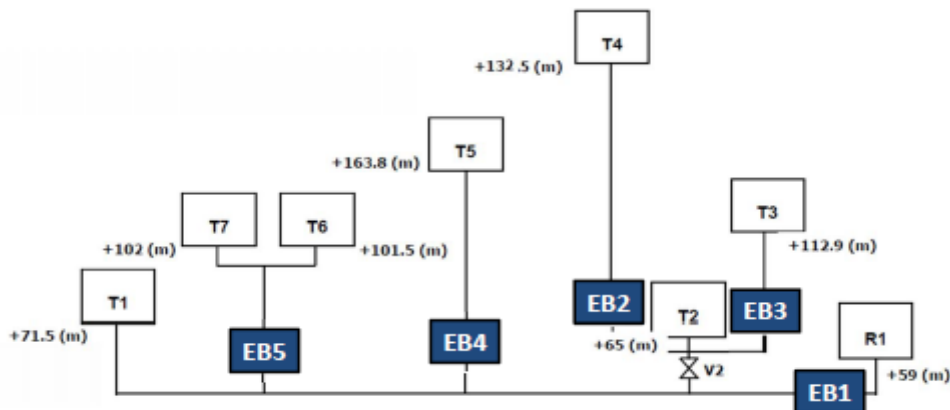


Figura 3. Líneas de caudal para los Sectores de la red

3.2.1 SECTOR 2

El sector 2 como vimos en la Tabla 1 cuenta con un tanque y una estación de bombeo conformada por dos bombas, de las cuales tan solo una está en funcionamiento. En cuanto al llenado de los tanques, BATADAL suministró un archivo de la red completa en EPANET, de donde se obtiene la geometría y con esto su volumen hora a hora a partir del cambio de nivel de agua que proporciona el conjunto de datos. Ese volumen por lo tanto puede salir del tanque para abastecer la ciudad o por lo contrario solo puede aumentar para llenar el tanque y abastecer en periodos posteriores.

El cambio de volumen se calcula a partir de la siguiente ecuación:

$$\Delta V_{(t)} = \frac{A_3}{3.6} [Z_{(t)} - Z_{(t+1)}]$$

Donde

$\Delta V_{(t)}$ = Cambio del volumen del tanque en el instante t

A_3 : Área del tanque 3

$Z_{(t)}$ = Nivel del agua en el tanque en el instante t

$Z_{(t+1)}$ = Nivel del agua en el tanque en el instante $t+1$

En cuanto a las bombas, se cuenta con datos de caudal bombeado de cada una de ellas, por lo que se propone tomar un valor medio en cada instante de tiempo y si existen arranques o paradas de las bombas, entonces se tomara el valor propio de esa hora. Entendiendo esto, el caudal total que inyecta la estación de bombeo será por lo tanto la suma de cada una de las bombas que esté funcionando en ese instante.

Con esta información se puede calcular ya el consumo de los usuarios a excepción de los instantes de arranques y paradas de las bombas. El resultado del consumo de los usuarios de este sector será por lo tanto la suma del caudal bombeado y la diferencia del volumen de agua en el tanque.

$$Q_{c2} = Q_b + \Delta V_T$$

Donde:

Q_{c2} : Caudal de consumo del sector 2

Q_b : Caudal bombeado por la estación 2

ΔV_T : Diferencia de nivel del tanque

Si se calculara el valor del caudal consumido de esta manera para los arranques y paradas de las bombas, se obtendrían valores negativos del consumo por lo que se plantea una alternativa diferente, a partir del cálculo del tiempo de encendido y paro de las bombas. Tomando como ejemplo la primera parada que sucede entre la hora 4 y 5 como vemos en la **Figura 4**, las bombas están encendidas hasta la hora 4 y llenan el tanque, sin embargo, en la hora 5 las bombas se han detenido y empieza por lo tanto el tanque a vaciarse y con eso el nivel del mismo a disminuir. En consecuencia, el tiempo de paro se ha calculado extrapolando las dos líneas de llenado y vaciado, dando lugar a un punto de intersección que en este caso será a las 4.12 horas.

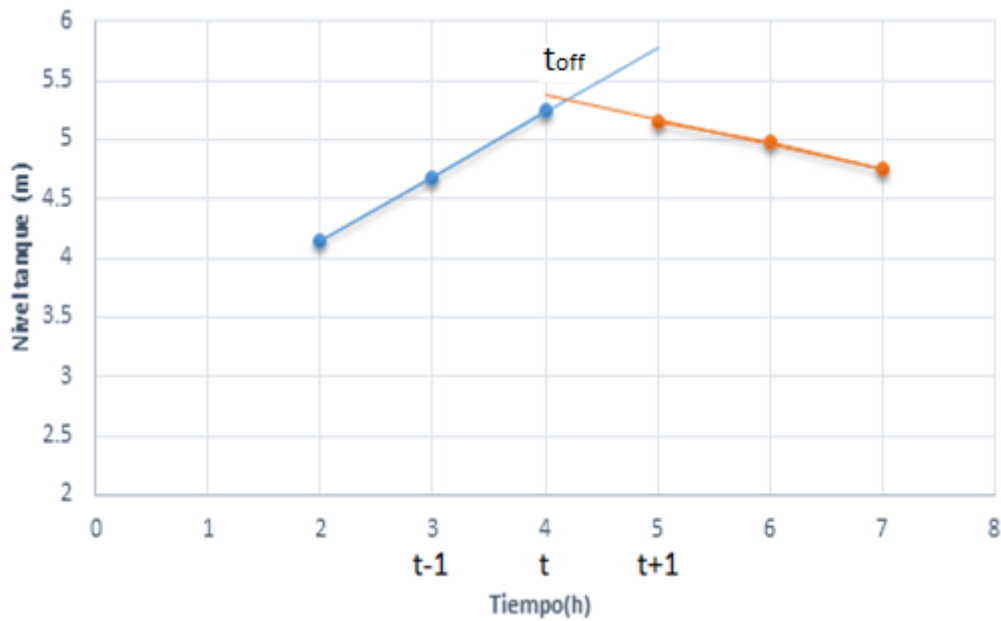


Figura 4. Nivel del Tanque 3 asociado al paro de la bomba.

Con el tiempo de paro de la bomba conocido, se aplica el concepto de balance de caudal para determinar el caudal consumido a través de la siguiente fórmula:

$$Q_{c2_{\text{paro}}} = \frac{A_3}{3.6} [Z_{(t)} - Z_{(t+1)}] + Q_{b(t)} \cdot t_{\text{off}}$$

Donde:

$Q_{c2_{\text{paro}}}$ = Caudal consumido en el paro de las bombas del Sector 2

A_3 : Área del tanque 3

$Z_{(t)}$ = Nivel del agua en el tanque en el instante t

$Z_{(t+1)}$ = Nivel del agua en el tanque en el instante $t+1$

$Q_{b(t)}$ = Caudal bombeado en el instante t

t_{off} = tiempo de apagado

Ahora si la estación de bombeo presenta un arranque como sucede entre la hora 11 y 12 de la **Figura 5**, donde el nivel del tanque empieza a aumentar, entonces se extrapola de la misma manera que en el paro de la bomba y se calcula el punto de igualdad para las dos rectas.

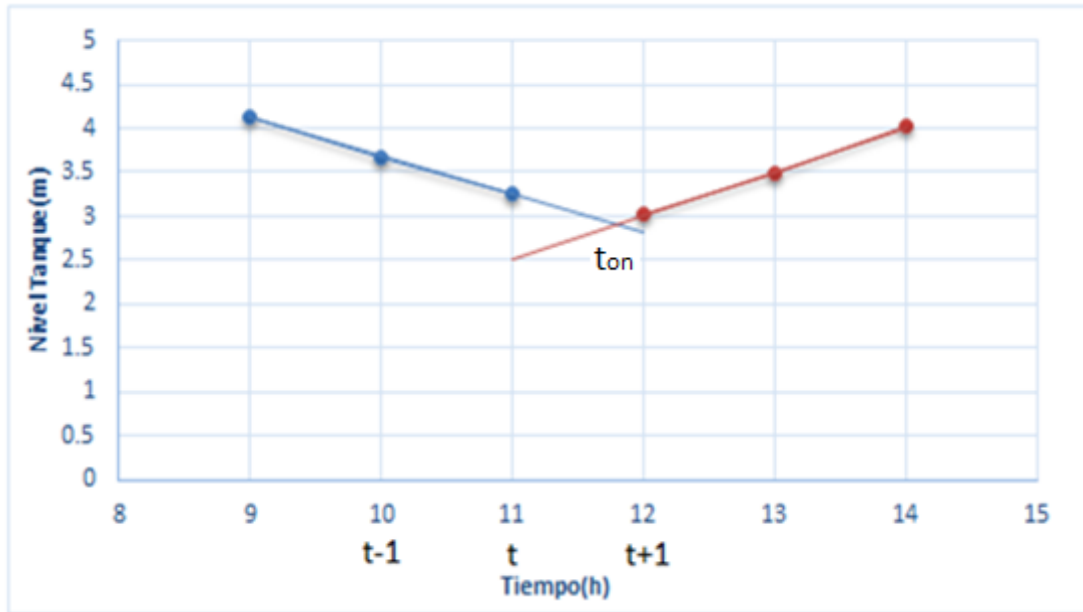


Figura 5. Nivel del Tanque 3 asociado al arranque de la bomba.

Para este ejemplo la hora de arranque de la bomba será 11.83 h y el caudal consumido se calculará por lo tanto a partir de:

$$Q_{c2_{arranque}} = \frac{A_3}{3.6} [Z_{(t)} - Z_{(t+1)}] + Q_{b(t+1)} \cdot (1 - t_{on})$$

Donde:

$Q_{c2_{paro}}$ = Caudal consumido en el arranque de las bombas del Sector 2

A_3 : Área del tanque 3

$Z_{(t)}$ = Nivel del agua en el tanque en el instante t

$Z_{(t+1)}$ = Nivel del agua en el tanque en el instante $t+1$

$Q_{b(t+1)}$ = Caudal bombeado en el instante $t+1$

t_{on} = tiempo de encendido

Con el caudal de consumo en la totalidad de los instantes, se procedió a calcular el patrón de consumo dividiendo el valor de la demanda de los usuarios en cada instante, entre la media de todos los valores calculados. De esta manera se obtuvo un patrón diario y se representó a través de una curva de modulación mostrada en la **Figura 6**.

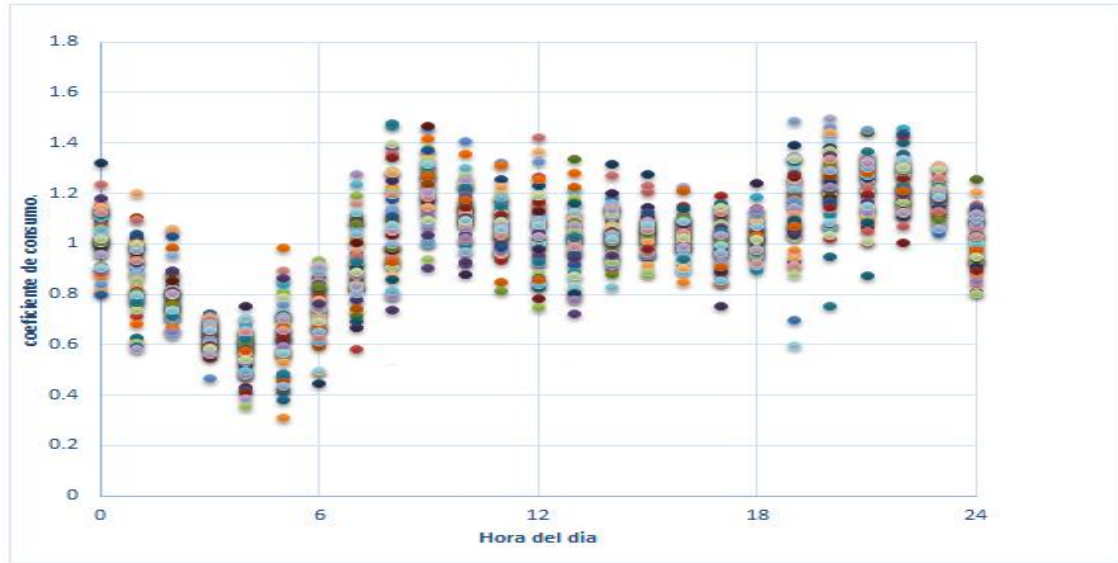


Figura 6. Curva de Modulación Sector 2

3.2.2 SECTOR 3

Este sector está conformado por dos bombas y un tanque, por lo que el caudal de las bombas y la variación del tanque se calculará de la misma manera que en el sector 2, sin embargo, para los arranques o paradas no se podrá realizar la extrapolación y de este modo calcular el caudal consumido, debido a que existen instantes donde se presentan estos dos casos uno tras otro como se muestra en la **Figura 7**

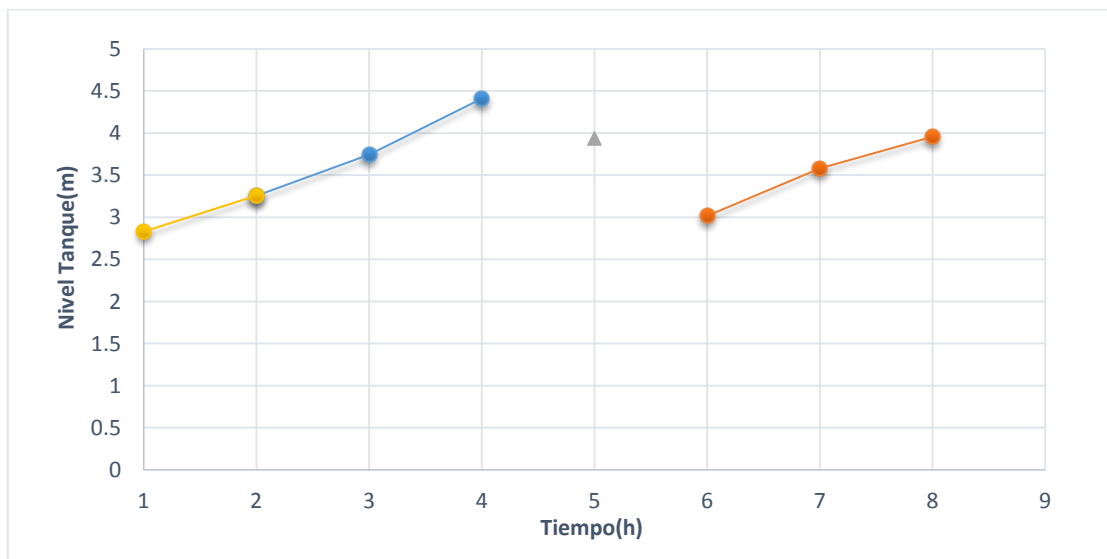


Figura 7. Nivel del Tanque 4 asociado a los arranques y paradas de la bomba.

Los datos afirman que en la hora 5 las bombas están paradas, de modo que en un momento comprendido entre la hora 4 -5 la bomba paró y entre la hora 5-6 la bomba nuevamente se puso en marcha. En consideración, para calcular el caudal consumido en estos dos intervalos, se contará solo con los datos aportados y se deducirá a partir de la media de los valores más próximos conocidos. Por ejemplo, para el instante $t=5$ el caudal consumido será la media del caudal consumido en $t=4$ y $t=6$.

Por último, se ha calculado también el caudal bombeado de igual forma que en el Sector 2 y la media de todos los valores de caudal consumido, para tener de así el coeficiente de consumo dividiendo toda la serie calculada por el valor promedio encontrado. A partir de esto se encontró la siguiente curva de modulación para el Sector 3 presentada a continuación:

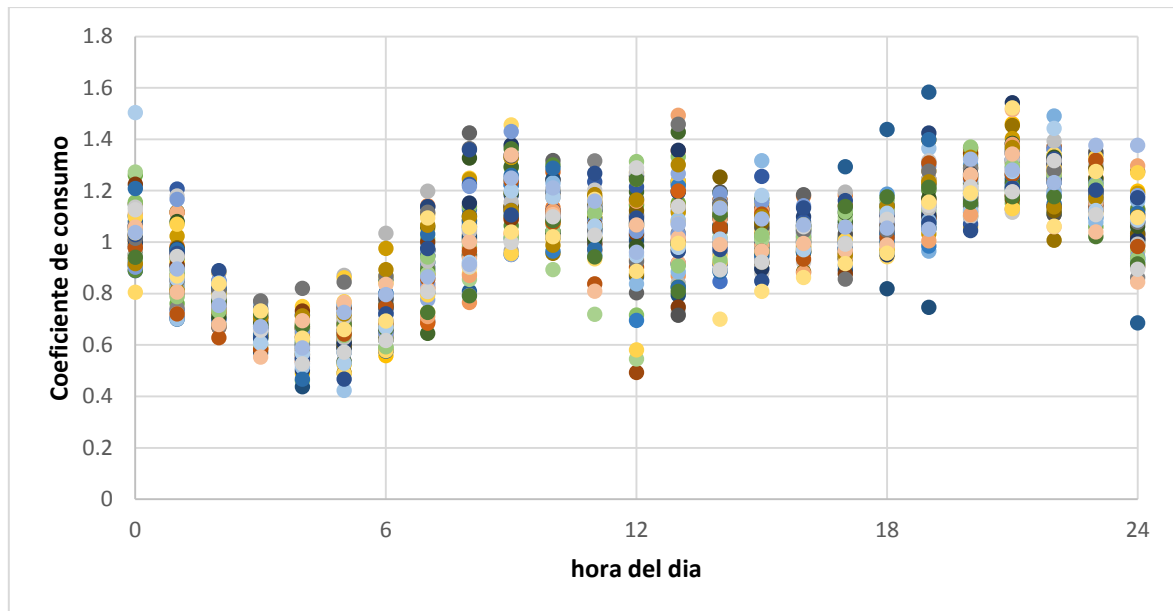


Figura 8. Curva de Modulación Sector 3

3.2.3 SECTOR 4

Este sector está conformado por un tanque y una estación de bombeo con 2 bombas, de las cuales solo una está en funcionamiento; particularmente el sector 4 tiene un comportamiento similar al del Sector 2, como vemos en la **Figura 9** donde existen periodos prolongados de funcionamiento y de parada de las bombas, por lo que se ha procedido de la misma manera, extrapolando y encontrando un punto de encendido y arranque de las bombas determinando así el caudal consumido como en el Sector 2.

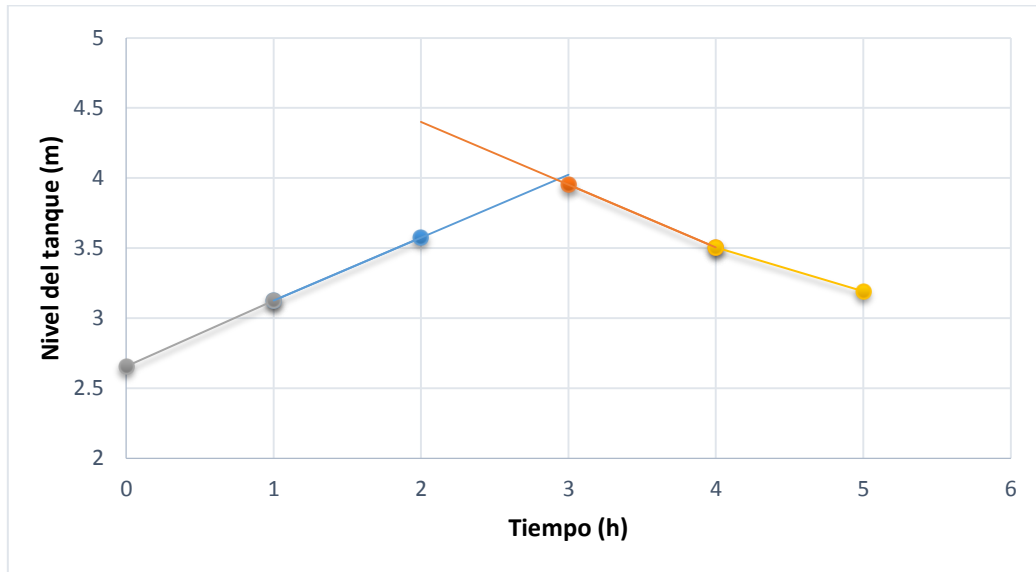


Figura 9. Nivel del Tanque 5 asociado al arranque de la bomba.

Con los resultados obtenidos se logró obtener un patrón de consumo para el sector 4 representado a través de una curva de modulación como se muestra a continuación:

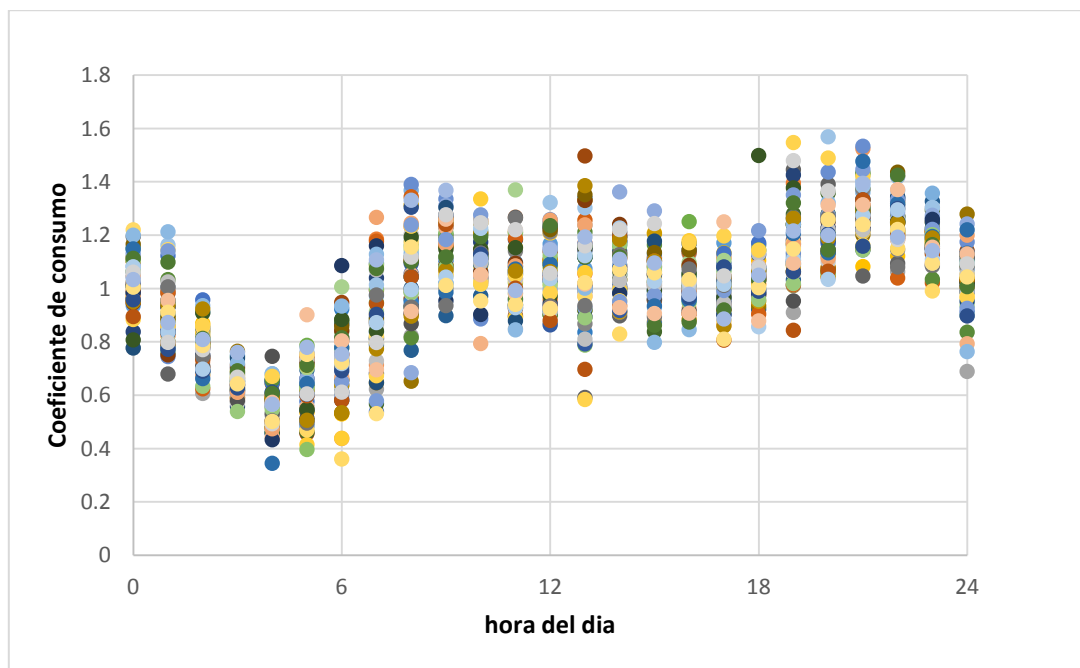


Figura 10. Curva de Modulación Sector 4

3.2.4 SECTOR 5

Este Sector, a diferencia de los anteriores, lo integra dos tanques y una estación con dos bombas. Por eso, para calcular la variación total del volumen, se han de sumar las variaciones individuales de cada tanque como se muestra en la siguiente ecuación:

$$\Delta V_T = \Delta V_6 + \Delta V_7$$
$$\Delta V_T = \left(\frac{A_6}{3.6} [Z_{(t)} - Z_{(t+1)}] \right)_{T6} + \left(\frac{A_7}{3.6} [Z_{(t)} - Z_{(t+1)}] \right)_{T7}$$

Donde:

A_6, A_7 = Área del tanque 6 y 7 respectivamente.

$Z_{(t)}$: Nivel del tanque respectivo en el instante t

$Z_{(t+1)}$: Nivel del tanque respectivo en el instante $t+1$

En cuanto al bombeo se ha de calcular igual que en los otros Sectores, no obstante, este sector tiene un comportamiento muy particular puesto que existen combinaciones de arranques y paradas que no se parecen a los sectores anteriores, como lo podemos visualizar en la siguiente figura:

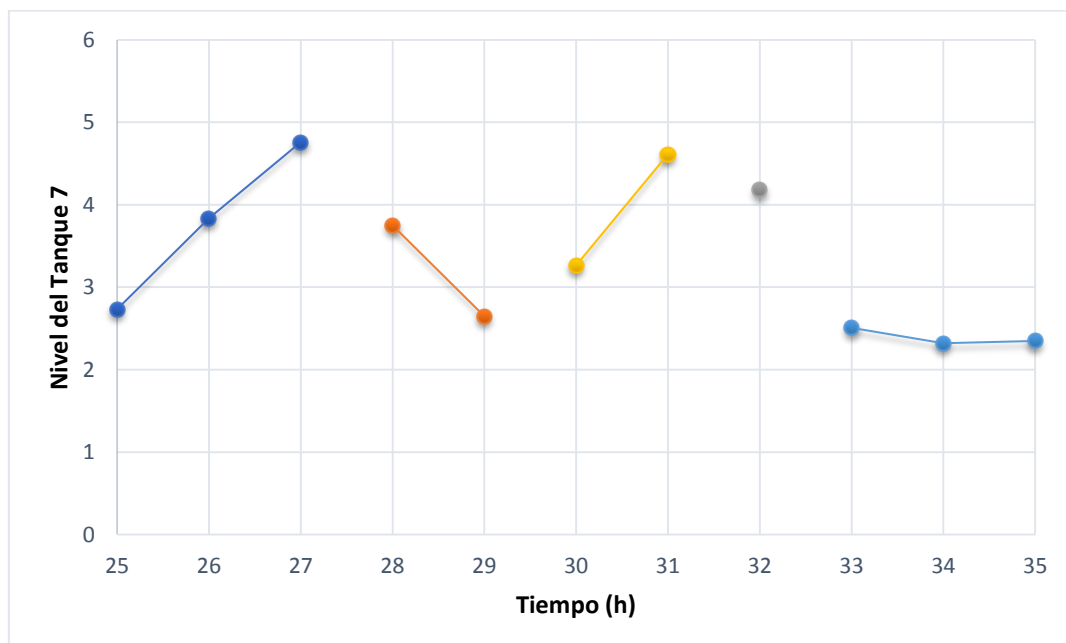


Figura 11. Nivel del Tanque 7 asociado a los arranques y paradas de la bomba.

Es decir que entre la hora 27-28 las bombas se detuvieron y se mantuvieron así hasta la hora 29, luego entre la hora 29-30 las bombas se encendieron y siguieron así hasta la hora 31, sin embargo, entre la hora 31-32 nuevamente se detuvieron y una hora después entre la hora 32-33 la bomba se enciende y continua así al final por algunas horas.

Este comportamiento se encontró en toda la serie de datos, por lo cual se optó al igual que en el Sector 3 por la media de los valores próximos conocidos, recordando que, si se hace con el procedimiento normal, sumando la variación del tanque más el caudal de las bombas, el caudal consumido resultaría negativo. Calculando para este ejemplo el consumo en la hora 32, sería entonces la media del valor del consumo en la hora 31 y 33. Con esta información, se ha podido calcular el patrón de consumo para este sector, dando como resultado la siguiente curva de modulación:

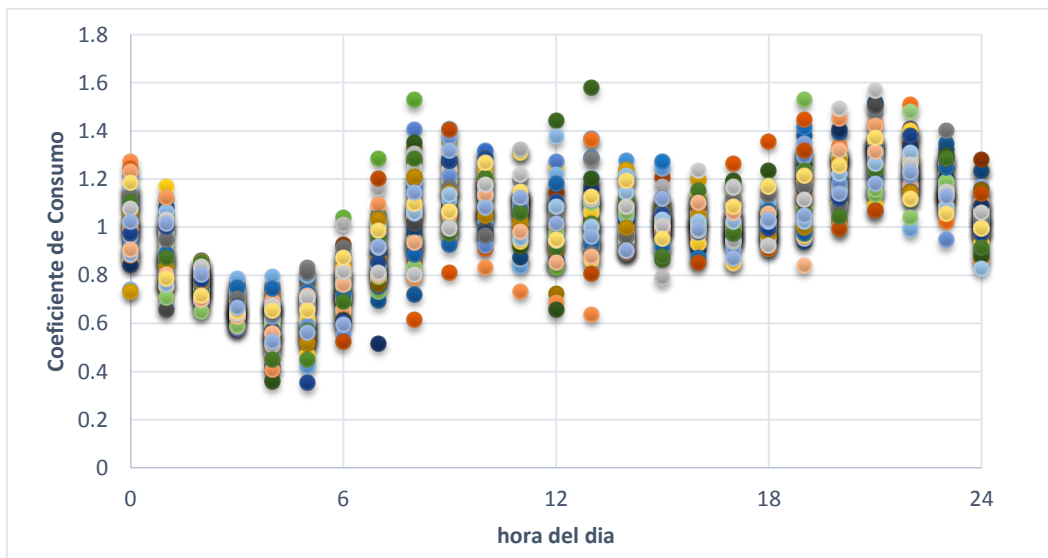


Figura 12. Curva de Modulación Sector 5

3.2.5 SECTOR 1

Como ya se había comentado antes, se ha dejado el cálculo de este sector al final debido a que todos los demás sectores dependen de él, es decir, que el agua que llega a cada uno de ellos pasa primeramente por este sector. El Sector 1 está formado por 2 tanques y una estación de bombeo con 3 bombas, de las cuales tan solo funcionan dos de ellas.

El cálculo del volumen de los tanques se realiza igual que en el Sector 4, sumando las variaciones de nivel de cada uno de ellos; respecto a las bombas, como se ha mencionado anteriormente, se ha

de sumar el caudal de cada una, sin embargo, del Sector 1 se abastecen los demás sectores como se ve en la **Figura 13**.

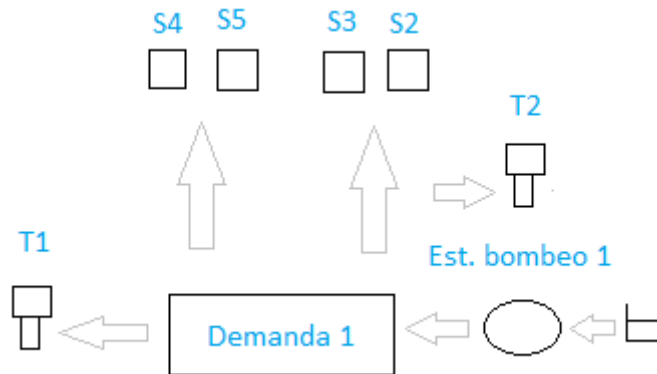


Figura 13. Flujo del caudal del Sector 1 a los demás Sectores

En consecuencia, el consumo del Sector 1, no solo depende de los tanques y del caudal bombeado por su estación correspondiente, sino también por los bombeos que en ese instante los otros sectores estén ejecutando, por lo cual el balance del caudal consumido para el Sector 1 será:

$$Q_{c1} = Q_b + \Delta V_T - \sum Q_{sec} ,$$

$$Q_b = (Q_{b1} + Q_{b2})$$

$$\Delta V_T = \left(\frac{A_1}{3.6} [Z_{(t)} - Z_{(t+1)}] \right)_{T1} + \left(\frac{A_2}{3.6} [Z_{(t)} - Z_{(t+1)}] \right)_{T2}$$

$$\sum Q_{sec} = Q(t)_{S2} + Q(t)_{S3} + Q(t)_{S4} + Q(t)_{S5}$$

Donde:

Q_b : Caudal bombeado en el sector 1 por las bombas PU1 y PU2

ΔV_T : Variación en los tanques T1 y T2

$\sum Q_{sec}$: Sumatoria de todos los caudales bombeados en los Sectores 2,3,4 y 5.

Q_{c1} : Caudal consumido en el sector 1

En cuanto a los arranques y paradas de las 2 bombas, este Sector tiene un comportamiento particular debido a que los dos tanques funcionan de forma distinta como se muestra en la Figura 13. El Tanque 1 funciona como el Tanque 3 del Sector 2, el cual se ha podido extrapolar y determinar el tiempo de encendido y apagado de las bombas, sin embargo, el Tanque 2 no presenta un comportamiento del que se pueda extrapolar, por lo cual se asumirá que los dos tanques se comportaran de la misma forma, en este caso como lo hace el Tanque 1.

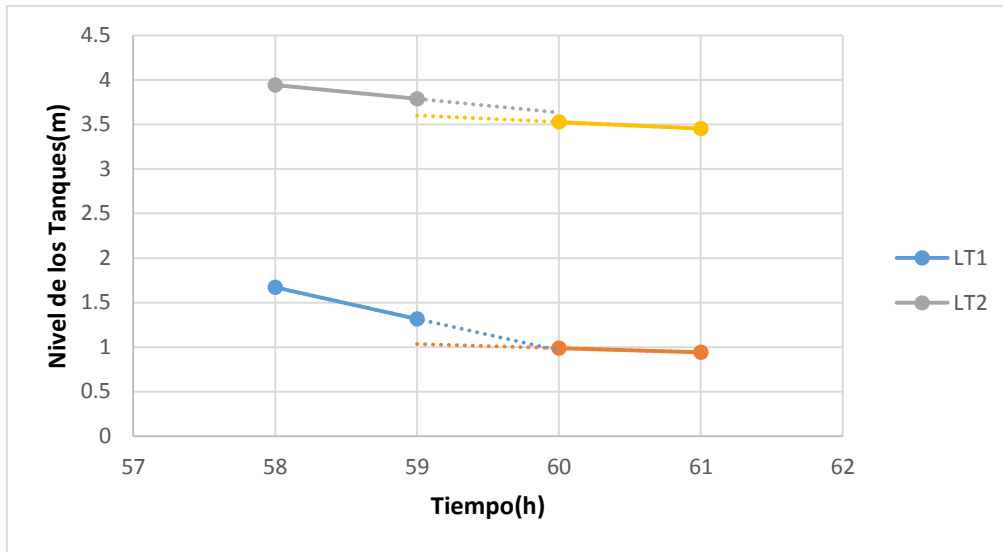


Figura 14. Nivel de los Tanques 1 y 2 asociado a los arranques de las bombas.

Obteniendo así el tiempo de arranque y parada de las bombas, se procedió a calcular el caudal consumido en estos instantes, de la misma manera que en los otros sectores que se pudo extrapolar. Con esta información, el patrón de consumo del Sector 1 representado a través de una curva de modulación es la siguiente:

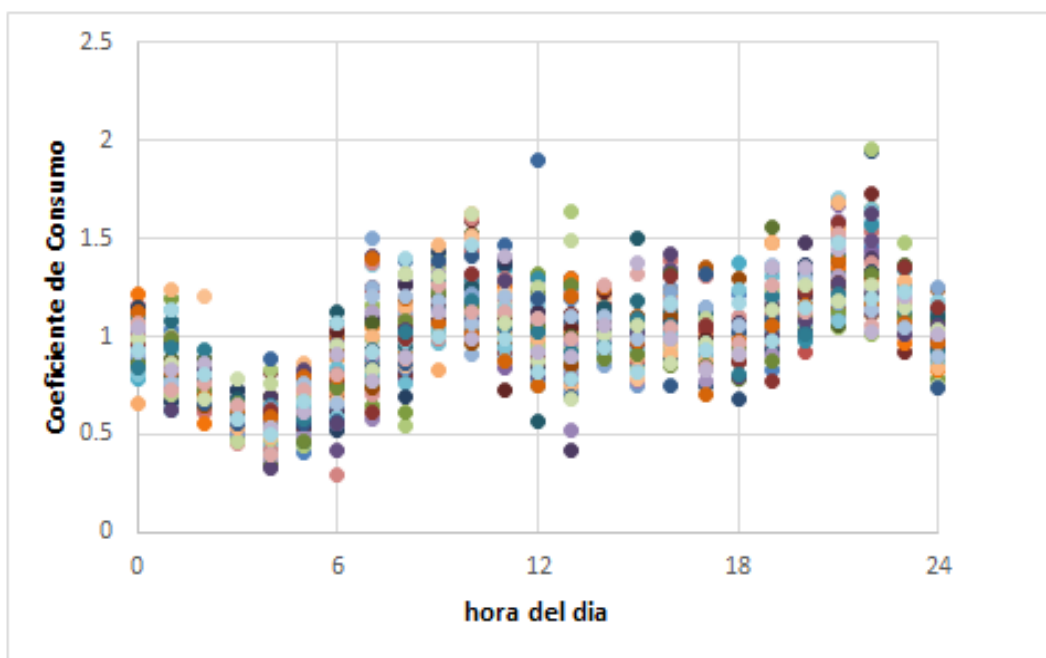


Figura 15. Curva de Modulación Sector 1.

3.3 DETERMINACION DE LAS CURVAS CARACTERISTICAS DE LAS BOMBAS

Después de haber calculado los patrones de consumo, es necesario interpretar de una forma más puntual lo que sucede en las estaciones de bombeo y determinar cuál es el funcionamiento correcto de estas y con eso, poder establecer un criterio respecto a las anomalías que se puedan presentar cuando agentes externos quieran manipular el proceso de bombeo.

Se tomará como base de cálculo el Sector 1 y se ha de aplicar la misma metodología en los demás sectores para determinar las curvas características correspondientes. Recordemos que este Sector posee 3 bombas, de las cuales solo 2 de ellas funcionan y se cuenta con los registros de caudales aportados por cada una de ellas, por lo cual solo bastaría calcular la altura de la bomba para cada uno de esos instantes. La descripción de la red la podemos observar en la siguiente figura:

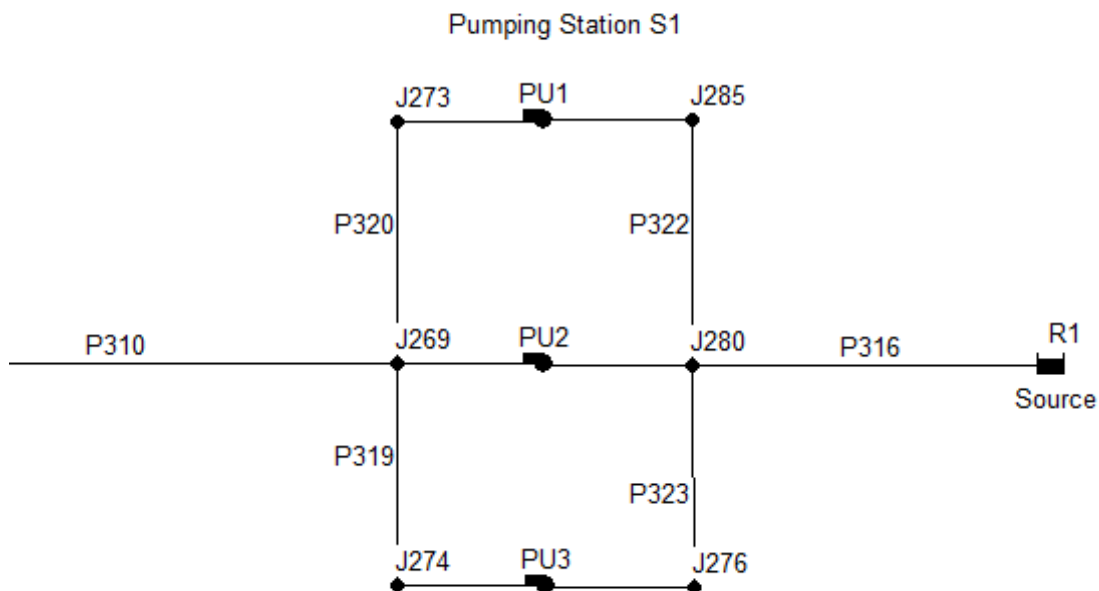


Figura 16. Estación de bombeo Sector 1.

Este sector cuenta con mediciones de la presión a la entrada y salida de la estación de bombeo, es decir en los puntos J280 y J269. Por otro lado, existe un tramo de tubería a la entrada y salida de la bomba PU1, por lo que será necesario calcular las pérdidas en estos tramos y así determinar al final la altura que genera cada una de las bombas. El esquema de la estación de bombeo se presenta en la Figura 16. Aplicando la ecuación de Bernoulli tenemos:

$$H_b = \left(Z_s + \frac{P_s}{\gamma} \right) - \left(Z_e + \frac{P_e}{\gamma} \right) + h_L$$

Donde:

Z_e : Cota a la entrada de la estación de bombeo, Punto J280

Z_s : Cota a la salida de la estación de bombeo, Punto J269

$\frac{P_e}{\gamma}$: Presión a la entrada de la estación de bombeo, Punto J280

$\frac{P_s}{\gamma}$: Presión a la salida de la estación de bombeo, Punto J269

h_L : Perdidas en la tubería de entrada y salida.

H_b : Altura de la bomba.

Con la altura de la bomba en cada instante de tiempo, se logró obtener las curvas características de la bomba PU1 y PU2 como se muestra en la **Figura 17** y **Figura 18**. Se evidencia por lo tanto un funcionamiento en paralelo en la bomba PU1 para caudales menores a 100 l/s, es decir, las dos bombas aportan el mismo caudal cuando la demanda de suministro aumenta.

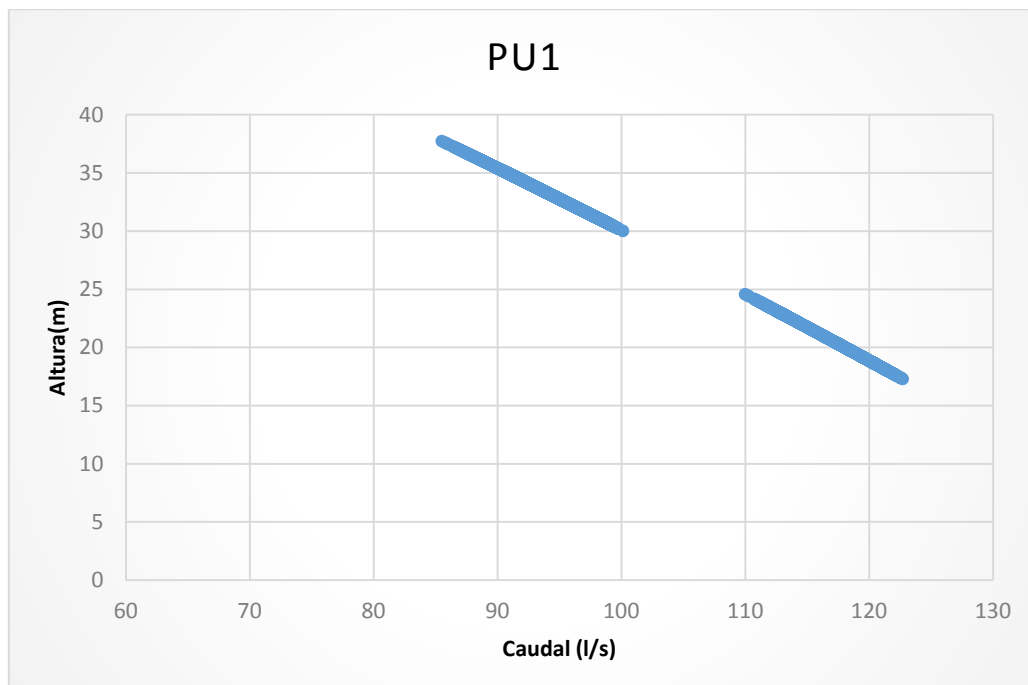


Figura 17. Curva característica de la bomba 1

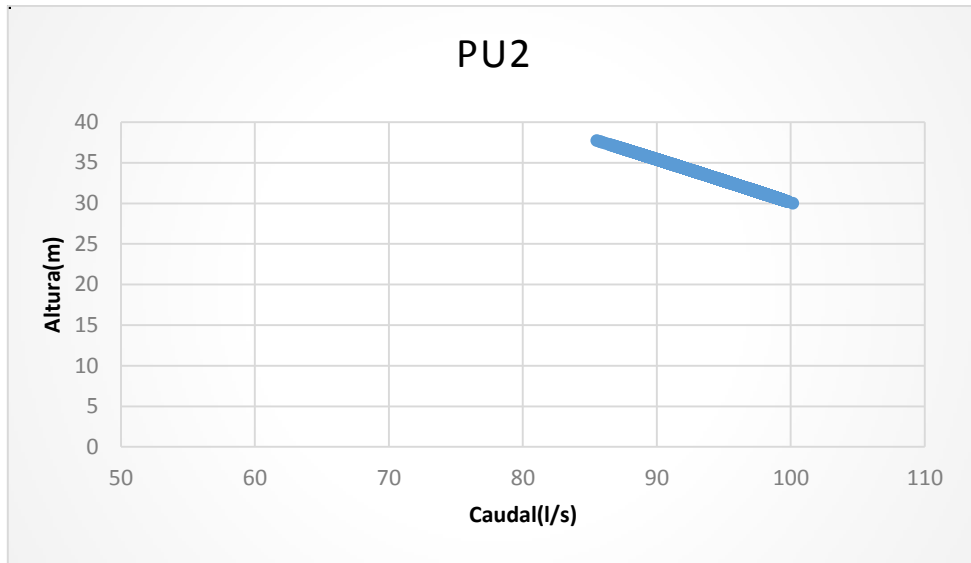


Figura 18. Curva característica de la bomba 2.

De la misma manera se obtuvo las curvas de las demás bombas que están en funcionamiento durante el desarrollo del dataset 1. En el caso de la bomba PU4, los caudales a bombear son más bajos y aunque en general la curva está bien definida, existe una pequeña fluctuación en algunos valores de modo que se tendrá posteriormente en cuenta este detalle cuando se compruebe los registros con anomalías. A continuación, se muestra la curva característica para esta bomba.

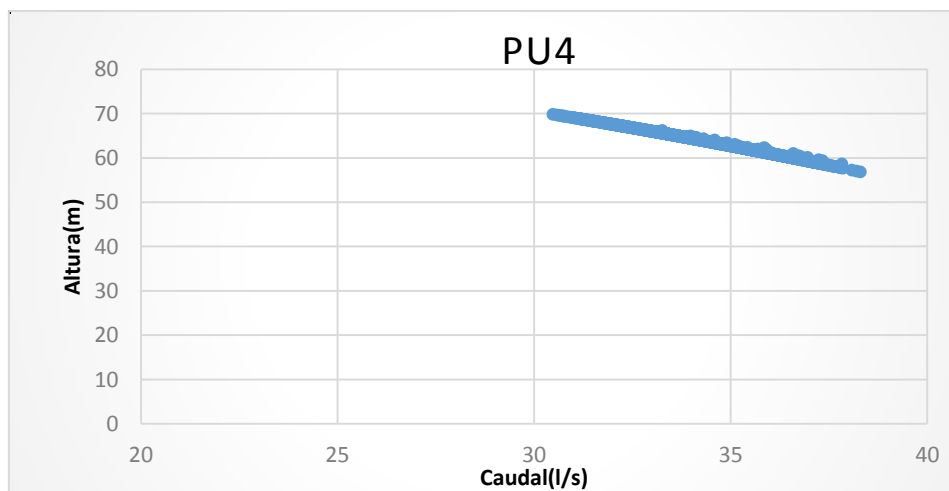


Figura 19. Curva característica de la bomba 4

En cuanto a la bomba PU6, los instantes en los cuales está encendida son muy pocos, sin embargo en la **Figura 20** se observa la forma de la curva característica a pesar de las pocas horas de funcionamiento.

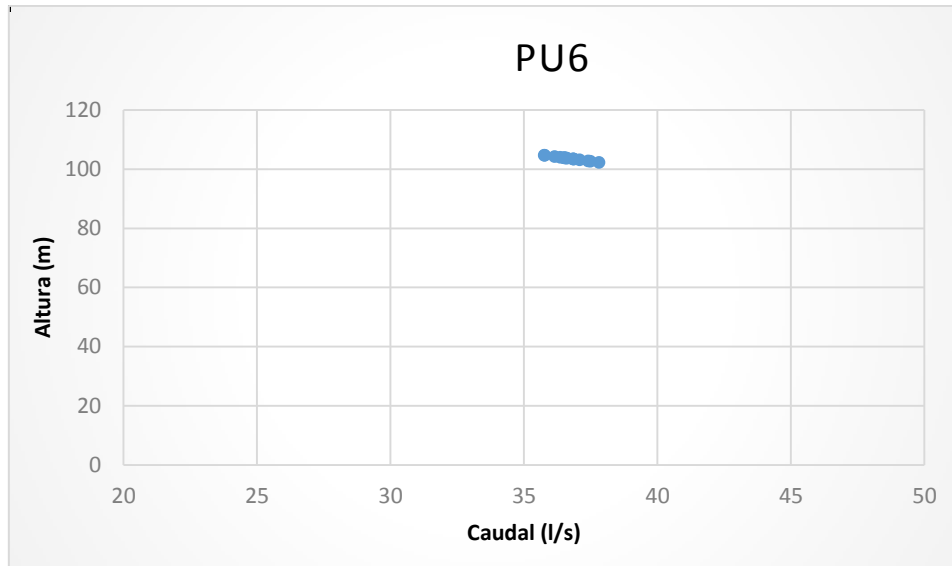


Figura 20. Curva característica de la bomba 6

De la misma manera, la bomba PU7 tiene un comportamiento parecido a la bomba PU1, donde tiene evidentemente un funcionamiento en paralelo para caudales menores a 40 l/s. De esta forma, los pocos instantes de funcionamiento de la bomba PU6 coinciden exactamente con los caudales en paralelo de la bomba PU7. La **Figura 21** describe este comportamiento.

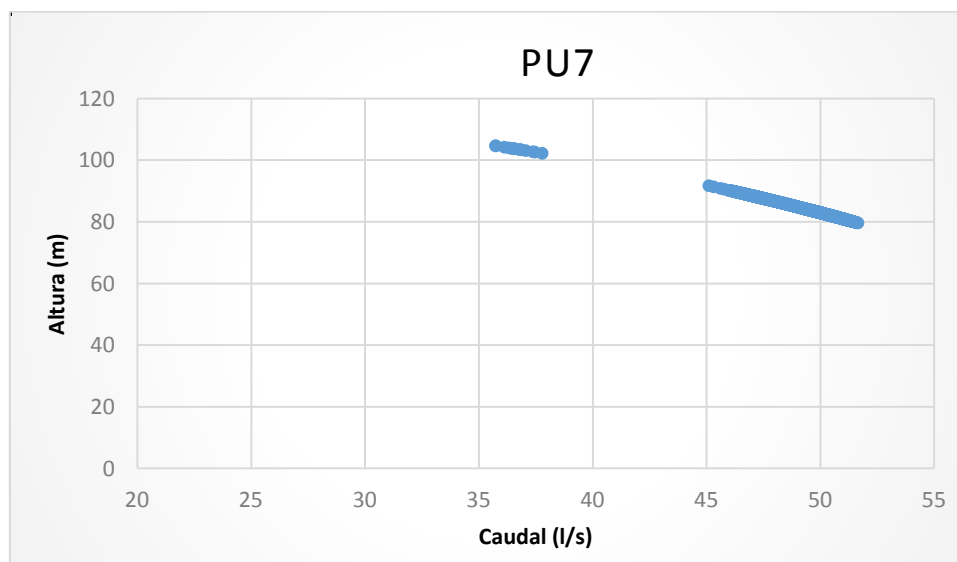


Figura 21. Curva característica de la bomba 7

Siguiendo con el cálculo de las curvas características, la bomba PU8 tiene un comportamiento habitual y maneja caudales bajos en comparación a las demás, sin embargo en la **Figura 22**, se muestra la curva bien definida para rangos entre los 32 y 39 l/s.

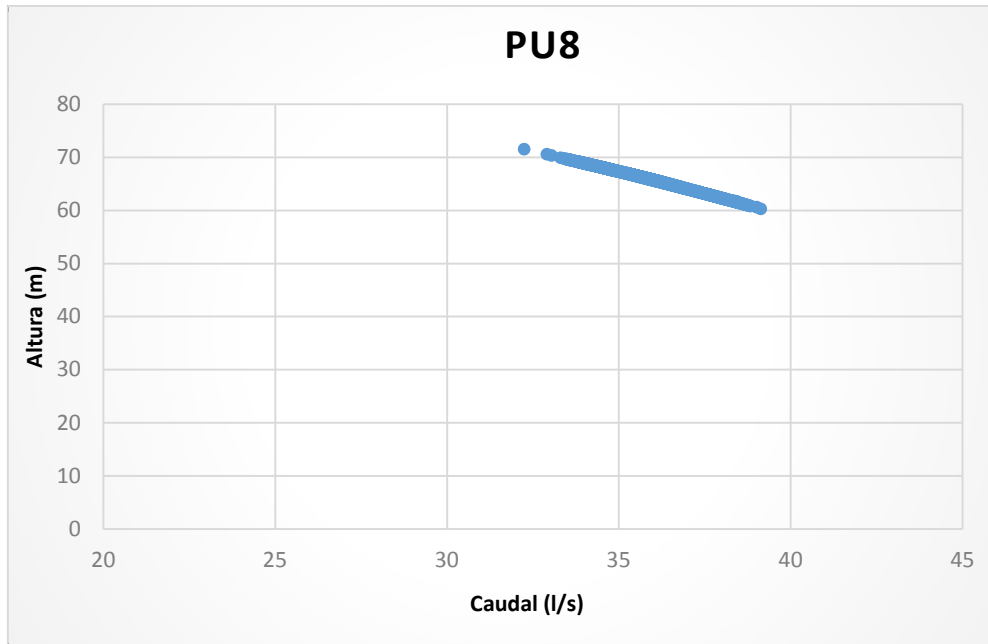


Figura 22. Curva característica de la bomba 8

En cuanto a las bombas PU10 y PU11 que pertenecen al Sector 5, las dos funcionan en paralelo en contados instantes, en este caso para caudales aproximadamente menores a 24 l/s. En las demás horas, la bomba PU10 funciona correctamente como se describe en la **Figura 23**.

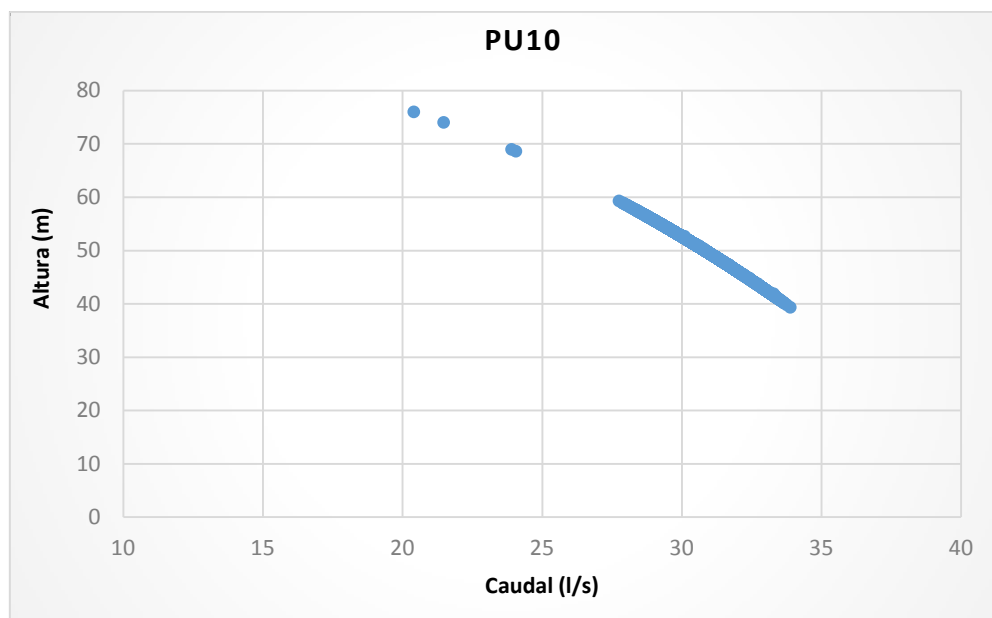


Figura 23. Curva característica de la bomba 10

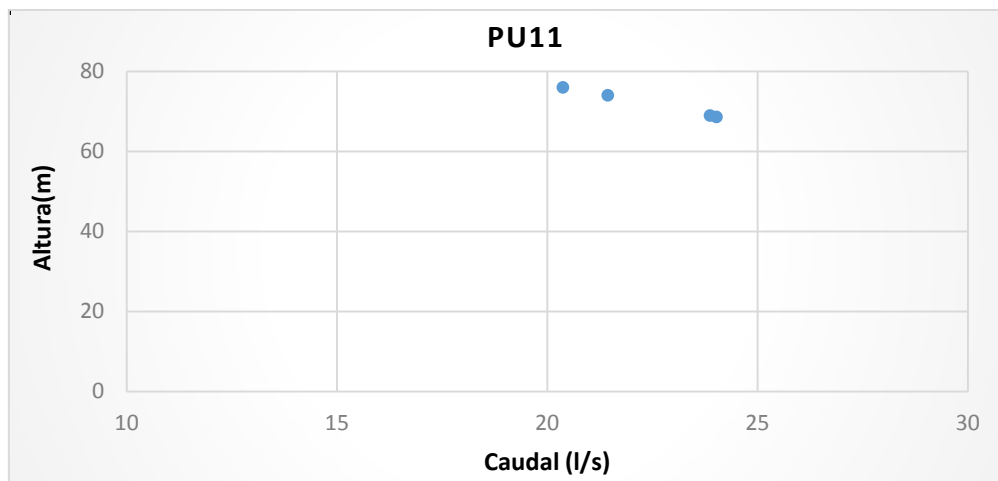


Figura 24. Curva característica de la bomba 11

Evidentemente, las curvas están bien definidas y muestran un comportamiento habitual de una bomba hidráulica, por ejemplo, en el Sector 3, a la que pertenecen la bomba PU6 y PU7, existe un rango de funcionamiento en paralelo para caudales entre 35 y 40 l/s. Lo mismo sucede para el Sector 1 donde se ve claramente el funcionamiento en paralelo de las bombas PU1 y PU2. También hay que mencionar que la bomba PU11 funciona en un par de momentos, en consecuencia, no se define bien la curva, sin embargo, coincide perfectamente con aquellos puntos alejados en la bomba PU10, donde las dos funcionan igualmente en paralelo.

Por otro lado, con la ayuda de la herramienta Línea de Tendencia de Excel, se calculó las ecuaciones que describen el funcionamiento de las curvas, que evidentemente se ajustan muy bien debido a la disposición correcta de los valores. A continuación, se muestran las ecuaciones para las curvas características de las bombas.

BOMBA	ECUACION
PU1	$H = -0.0009Q^2 - 0.3524Q + 74.788$
PU2	$H = -0.007Q^2 + 0.7971Q + 20.666$
PU4	$H = -0.0654Q^2 + 2.8439Q + 43.355$
PU6	$H = -0.025Q^2 + 0.6511Q + 113.33$
PU7	$H = -0.0292Q^2 + 0.9871Q + 106.73$
PU8	$H = -0.0706Q^2 + 3.4348Q + 33.638$
PU10	$H = -0.0742Q^2 + 1.326Q + 79.689$
PU11	$H = -0.0665Q^2 + 0.9301Q + 84.632$

Tabla 7. Ecuaciones de las curvas características en todas las bombas

3.4 DETERMINACION DE LOS CONTROLES.

El funcionamiento de las estaciones de bombeo está directamente relacionado al comportamiento de los tanques de almacenamiento, debido a que tanto el encendido como el apagado depende de los controles configurados a partir del rango del nivel del tanque. Es por eso que se vio la necesidad de analizar este comportamiento y estudiar la posibilidad de establecer un criterio que nos conduzca a la detección de fallos.

La implementación de controles en las estaciones de bombeo viene por la necesidad de controlar el caudal que se bombea y evitar desbordamientos o situaciones donde los tanques se encuentren totalmente vacíos. Por ejemplo, cuando el nivel del tanque ha descendido hasta cierto valor, las bombas inmediatamente se encienden, evitando así que el tanque quede sin reserva de agua. Por otro lado, si el nivel del tanque ha sobrepasado el nivel máximo impuesto por su geometría y su operatividad, también es necesario que las bombas se apaguen y se pueda evitar desbordamientos en estos elementos hidráulicos.

En este mismo sentido, se ha de analizar el comportamiento de las bombas frente al nivel del tanque, esperando que cuando las bombas estén encendidas, el nivel del tanque aumente, de forma que se pueda crear un criterio a partir del cual se identifique instantes que no cumplan esta condición y marcarlos como sospechosos. En la **Figura 25** se muestra el funcionamiento de las primeras horas del Tanque 4 y la estación de bombeo 3 conformada por las bombas PU6 y PU7.

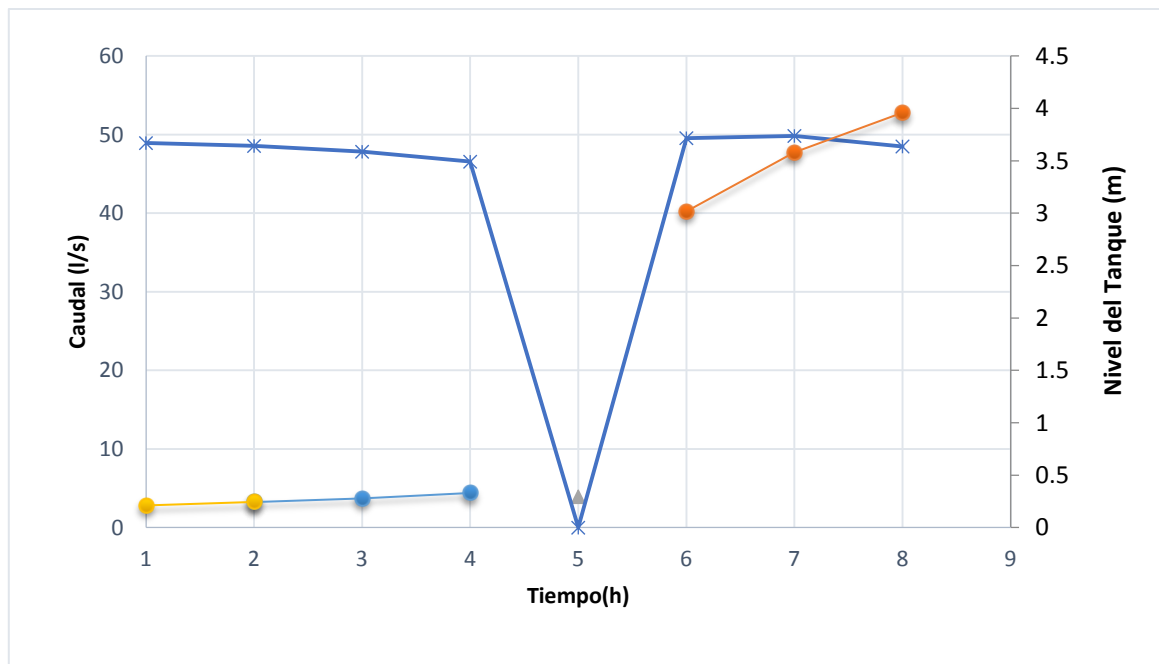


Figura 25. Comportamiento de las bombas respecto al nivel de los tanques.

Observando este comportamiento, se puede afirmar que, en las primeras cuatro horas, el tanque se está llenando puesto que su nivel aumenta, sin embargo, la bomba inyecta cada vez menos agua con el pasar de las horas. Luego, en la hora 5, el tanque se vacía mientras la bomba ha detenido su funcionamiento; por último en la hora 6, la bomba aparece nuevamente encendida y de la misma forma aumenta abruptamente el nivel del tanque.

Este funcionamiento se ha encontrado en todas las estaciones de bombeo, donde efectivamente se descarta una relación directa entre el llenado de los tanques y el arranque de las bombas, puesto que, como se vio en el ejemplo anterior, el tanque puede vaciarse y llenarse cuando la bomba está encendida, de manera que no se puede establecer un criterio a partir del cual se pueda asumir que el tanque solo puede llenarse cuando las bombas estén encendidas.

De este modo, el enfoque basado en los controles, queda totalmente descartado como una forma de detección y aunque en esta red particular no haya dado los resultados que se esperaba, muy seguramente es aplicable a otras redes que se ajusten a los requerimientos de este criterio.

3.5 ANALISIS DE LOS TANQUES

La red C-Town posee un sistema de regulación basado en la implementación de siete Tanques, los cuales controlan el arranque y parada de las estaciones de bombeo de acuerdo al nivel en que se encuentren, por eso para poder entender este funcionamiento, a continuación, se obtendrá a partir del dataset 1, comportamientos correctos respecto a los niveles de los tanques y así tener claridad sobre los niveles máximos o mínimos que pueden alcanzar antes de producirse desbordamientos que afecten al sistema.

Primeramente, el sector 1 cuenta con los Tanques 1 y 2, los cuales poseen las siguientes características hidráulicas:

TANQUE	AREA (m ²)	NIVEL MAX (m)	NIVEL MIN (m)
1	769.4	5	0
2	339.15	5.7	0

Tabla 8. Propiedades de los Tanques Sector 1

Donde se ve claramente que el Tanque 1 es aproximadamente el doble de tamaño que el Tanque 2, aunque sus niveles de operación son relativamente similares, por lo tanto, será necesario no solamente calcular los niveles de operación a partir del dataset 1, sino observar el comportamiento durante toda la simulación para poder observar la disposición de los valores en el tiempo. La **Figura 26** y **Figura 27** muestran este comportamiento.

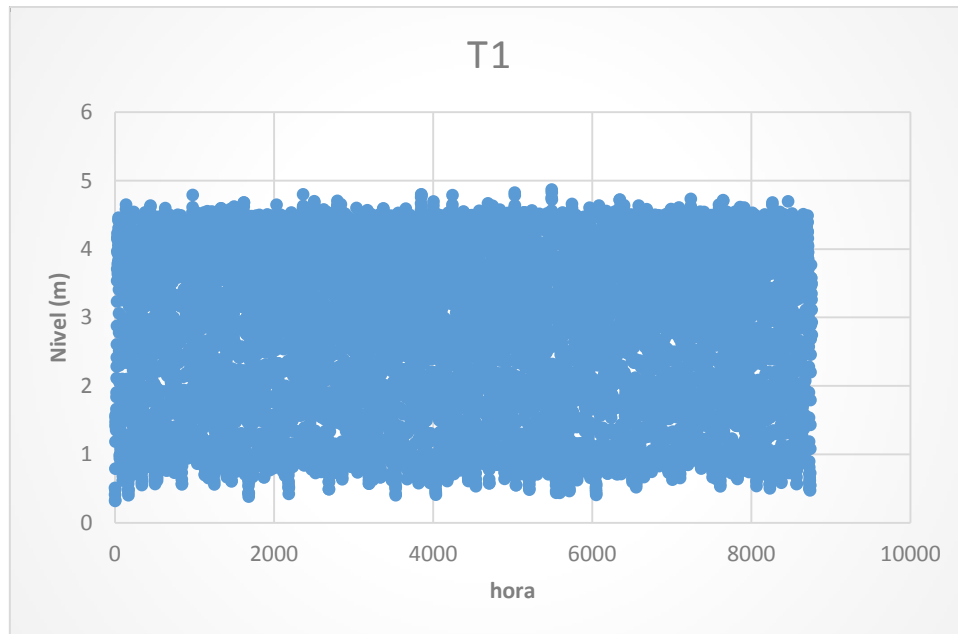


Figura 26. Niveles del Tanque 1 dataset 1.

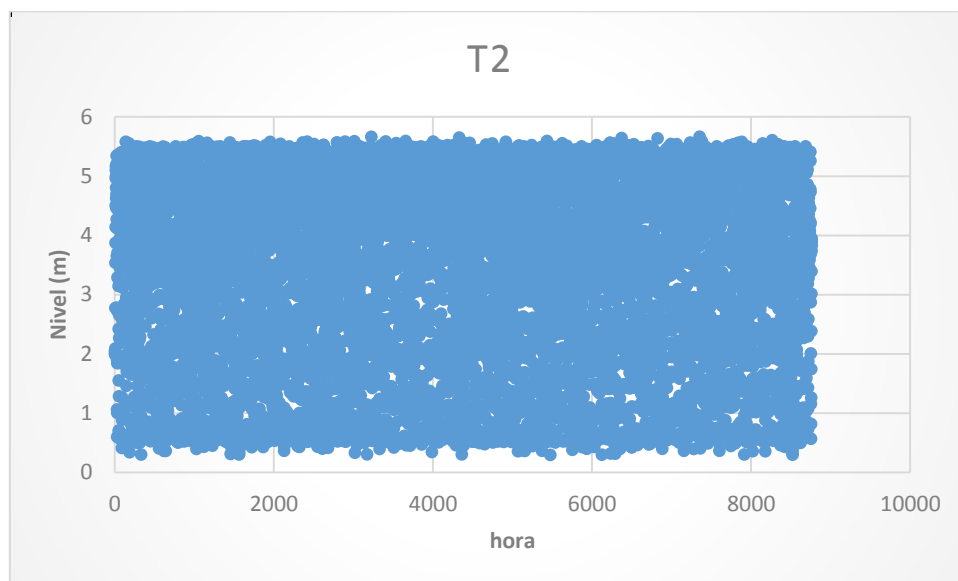


Figura 27. Nivel del Tanque 2 dataset 1.

Analizando estos valores podemos concluir que los niveles son bastante constantes en el Tanque 2 y mantiene un rango bastante definido, que en ningún momento presenta ocasiones de desbordamiento. En cuanto al Tanque 1, la disposición de los valores es un poco más fluctuante, pero se debe a la importancia de este Sector dentro de todo el funcionamiento de la red, puesto que el tanque no solamente está sometido a la variabilidad de los consumos del sector 1 sino también a los consumos y bombes de los demás sectores que precisamente se abastecen de este.

Para el Sector 2, el tanque correspondiente es el Tanque 3, el cual controla las bombas PU4 y PU5 y posee las siguientes características:

TANQUE	AREA (m2)	NIVEL MAX (m)	NIVEL MIN (m)
3	148.05	5.5	2.5

Tabla 9. Propiedades del Tanque Sector 2

Como se ve, el Tanque 3 tiene dimensiones mucho más pequeñas que los tanques anteriores y además posee un límite inferior de valor 2.5 m. En el mismo sentido, los valores de los niveles de los tanques durante toda la simulación se recogen en la siguiente figura:

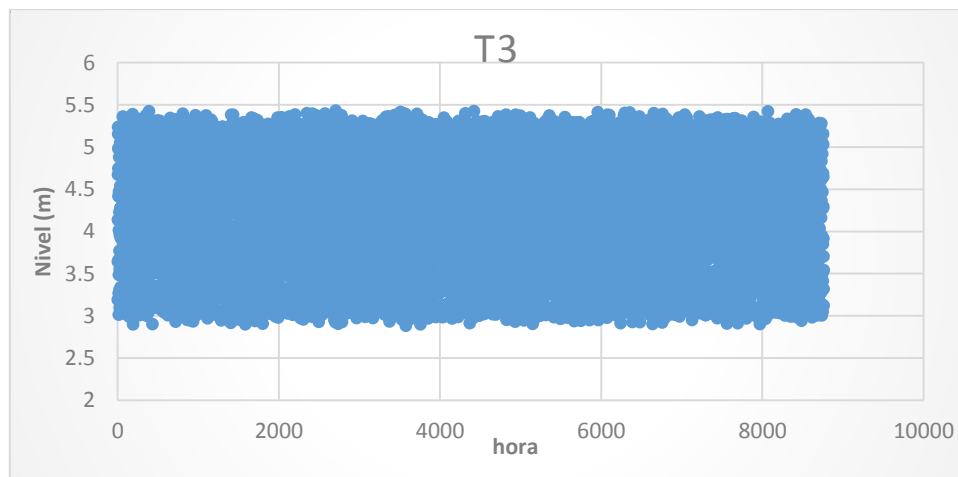


Figura 28. Nivel del Tanque 3 dataset 1

La **Figura 28** nos muestra un comportamiento relativamente constante, marcando claramente los límites superior e inferior definidos en la **Tabla 9**. Con estos valores se ha establecido el comportamiento normal para este Sector respecto al Tanque 3.

En cuanto al Sector 3, el Tanque 4 define el comportamiento de las bombas PU6 y PU7; además posee las siguientes características:

TANQUE	AREA (m2)	NIVEL MAX (m)	NIVEL MIN (m)
4	107.32	5	2

Tabla 10. Propiedades del Tanque 4 Sector 3

El Tanque 4 es dimensionalmente más pequeño que los anteriores y posee además límites inferiores y superiores. A continuación, se observa la variabilidad de los niveles del tanque durante toda la simulación de los datos:

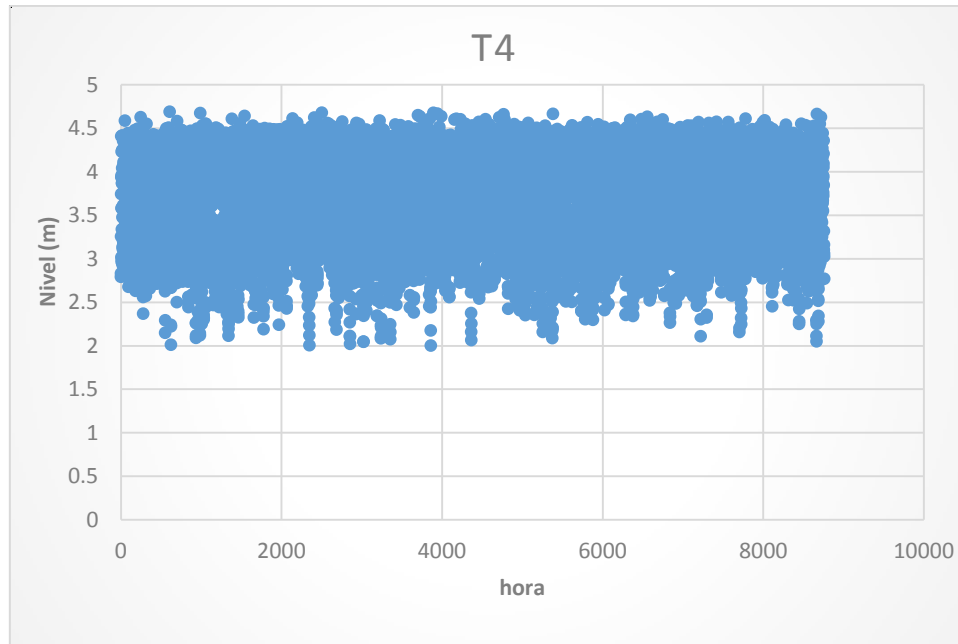


Figura 29. Nivel del Tanque 4 dataset 1

Como se observa en la **Figura 29**, los datos en el límite superior son relativamente constantes, sin embargo respecto al límite inferior, existe muchas fluctuaciones, pero con valores que no sobrepasan los valores límites de la **Tabla 10**.

El Sector 4 contiene al Tanque 5, el cual controla a su vez a las bombas PU8 y PU9 donde a su vez posee las siguientes características:

TANQUE	AREA (m2)	NIVEL MAX (m)	NIVEL MIN (m)
5	111.03	4.5	1.2

Tabla 11. Propiedades del Tanque 5 Sector 4

Este tanque es de dimensiones parecidas al Tanque 4 y posee un límite inferior y superior de funcionamiento. De igual forma es necesario analizar el comportamiento del tanque en todas las horas de registro del dataset1 a través de la siguiente figura:

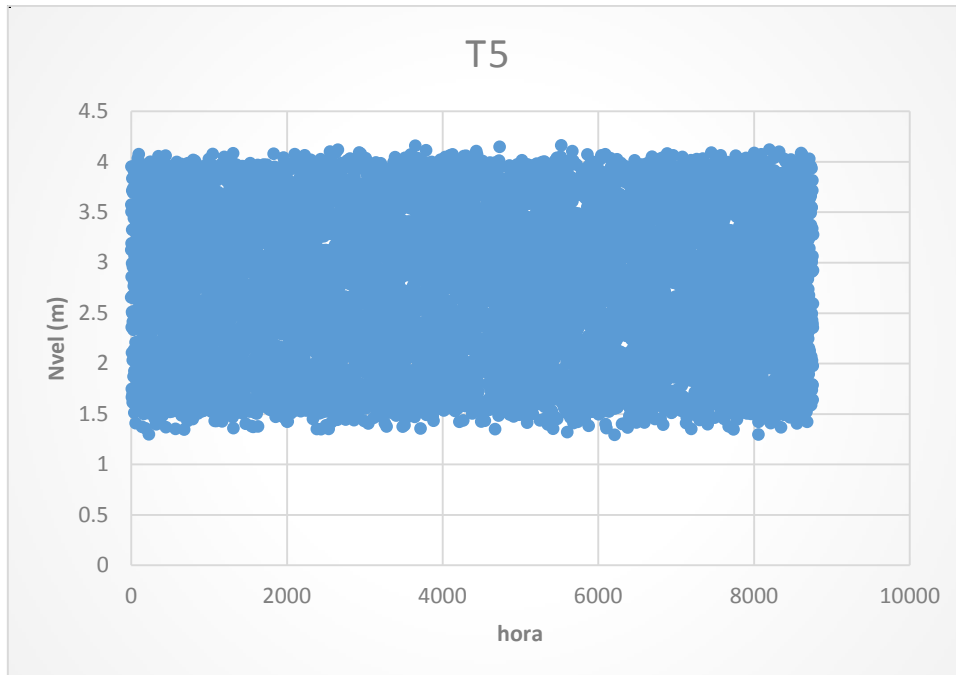


Figura 30. Nivel del Tanque 5 dataset 1

La **Figura 30** nos muestra una disposición continua en el Tanque 4, viéndose fácilmente los dos límites en los valores que coinciden precisamente con los valores de la **Tabla 11**, los cuales corroboran el comportamiento normal del Tanque.

En el último sector, esto es en el Sector 5, pertenecen los Tanques 6 y 7; los cuales a su vez controlan las bombas PU10 y PU11 y poseen las siguientes características:

TANQUE	AREA (m ²)	NIVEL MAX (m)	NIVEL MIN (m)
6	54.5	5.5	0
7	40	5	1

Tabla 12. Propiedades del Tanque 6 y 7 dataset 1

Como se comprueba en la **Tabla 12**, la dimensión de estos dos tanques son muy pequeñas en comparación a los demás, por otro lado el tanque 6 solo posee límite superior, mientras que el tanque 7 si presenta tanto límite superior como inferior. Entendido esto, se muestra a continuación la ubicación de los valores de los dos tanques a través de todas las horas de registro.

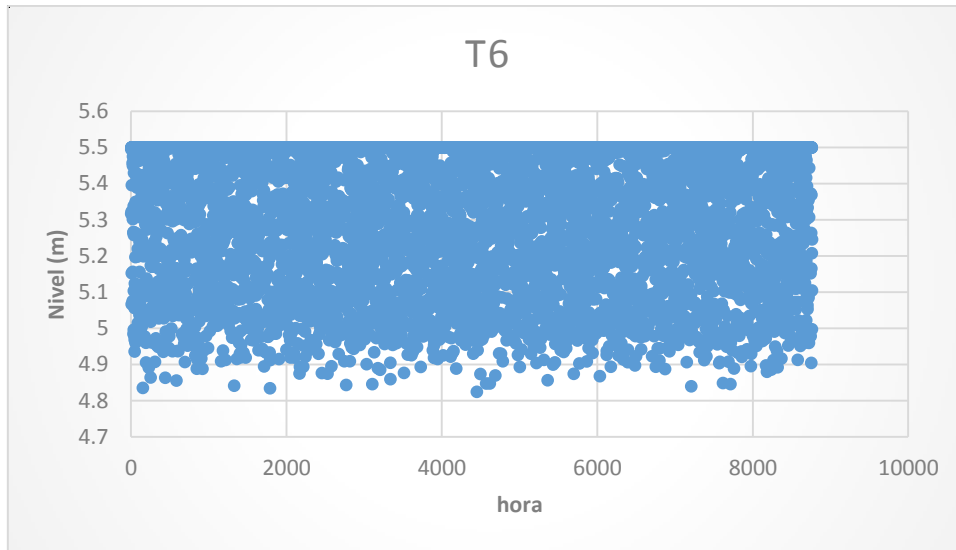


Figura 31. Nivel del Tanque 6 dataset 1

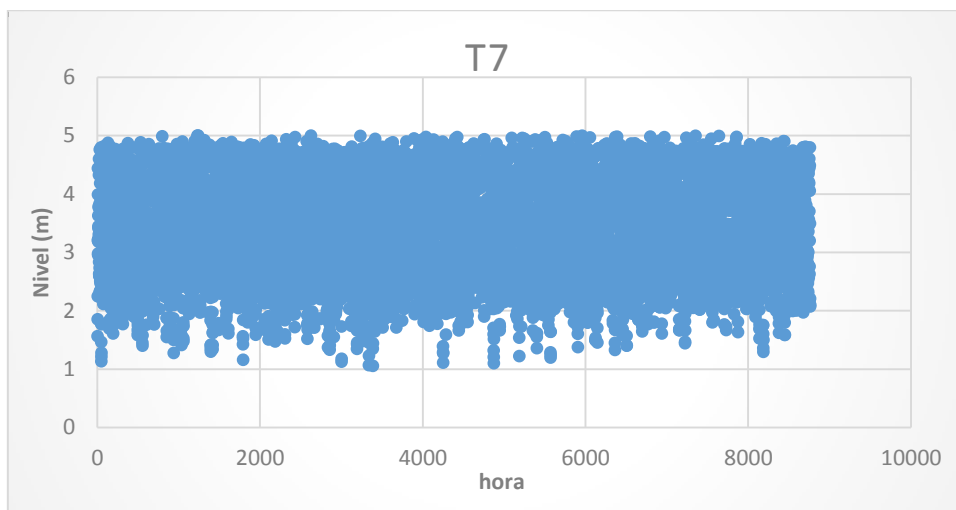


Figura 32. Nivel del Tanque 7 dataset 1

En la **Figura 31** se observa un funcionamiento continuo en cuanto al límite superior pero algún grado más de variabilidad sobre el límite inferior y esto es entendible debido a la pequeña geometría del Tanque. Por otro lado en la **Figura 32** existe aún más variabilidad que en el tanque anterior y como miramos primeramente, todo esto es causado por el tamaño pequeño del tanque.

Con estos valores analizados, se creará posteriormente un rango habitual y normal de funcionamiento en estos elementos hidráulicos, de tal manera que se pueda detectar ciertas anomalías con base a estos comportamientos encontrados.



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA

máster en ingeniería
hidráulica y medio ambiente
mihma

4. COMPROBACION DE REGISTROS.

Después de elaborar los criterios y enfoques del modelo con el fin de detectar un posible fallo en los elementos hidráulicos, se procede a usar directamente estos argumentos que consisten en el análisis de los consumos, la descripción correcta del funcionamiento de las bombas y la verificación de los estados de los tanques; a través del uso del segundo conjunto de datos o dataset 2, el cual ha sido brindado por la recopilación de información del sistema en los periodos donde efectivamente se llevaron a cabo anomalías.

Durante toda la comprobación de los criterios, se usará los datos encontrados en el dataset 1 y se compararan con los resultados ofrecidos por el modelo. Así se empezará primeramente con el análisis de los patrones de consumo respectivos, pasando por las estaciones de bombeo y por ultimo por el comportamiento de los tanques.

4.1 VALIDACION DE CONSUMOS.

BATADAL como se lo ha mencionado anteriormente, ha logrado obtener y proporcionar registros con anomalías en la red durante algunos meses, con la idea de plantear una metodología de detección capaz de ser verificada. Por eso, para llevar a cabo esta comprobación, se ha creado un criterio a partir del cual todos aquellos que no lo cumplan, serán datos sospechosos y posiblemente anomalías en el sistema.

El criterio está basado en el concepto de comparación de consumo, es decir, a partir del dataset 1, se buscará un rango lógico de comportamiento de los usuarios de la red y con eso separar aquellos registros que tengan un comportamiento inusual. El criterio de validación es el siguiente:

- Todas aquellas medidas que se salgan del valor máximo o mínimo encontrado en el dataset 1, serán clasificados como sospechosos.

4.1.1 VALIDACION DE CONSUMOS SECTOR 1

Después de elaborar los criterios de evaluación del modelo a partir del dataset 1, se procede a ingresar al modelo de detección los registros del dataset 2, los cuales contienen anomalías en varios periodos de tiempo. Por ejemplo, para este sector el cual incluye la bomba PU1 y PU2 se encontró la siguiente curva de modulación:

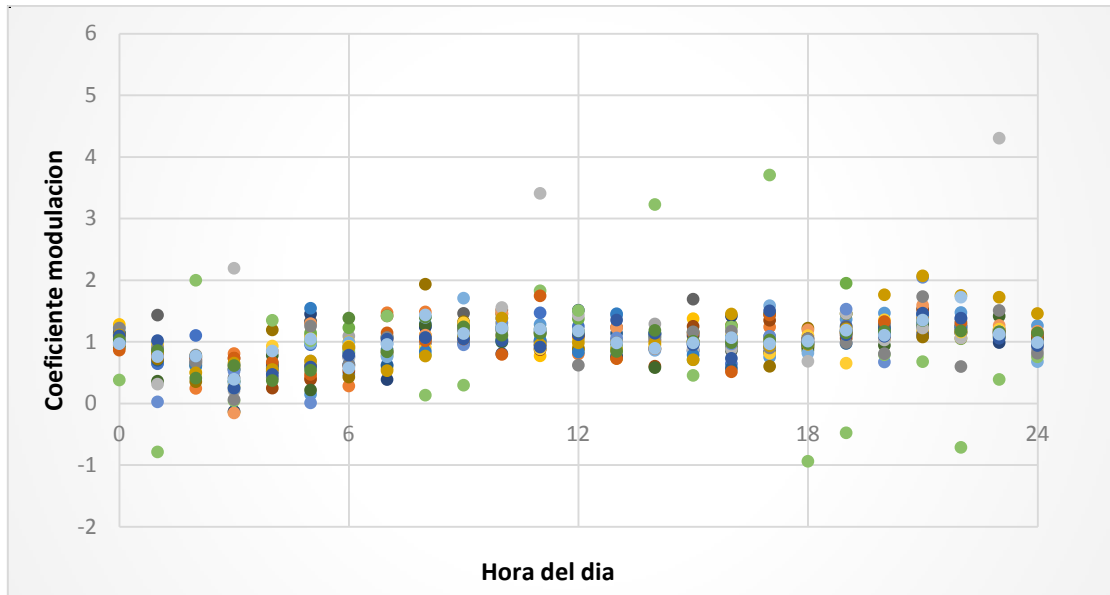


Figura 33. Curva de modulación Sector 1 dataset 2.

Observando la gráfica notamos que existen valores que no se ajustan a la tendencia de la curva y probablemente se traten de anomalías en el sistema, en el mismo sentido, la **Figura 34** muestra el análisis de continuo durante todo el proceso, mostrando de una mejor manera las anomalías existentes.

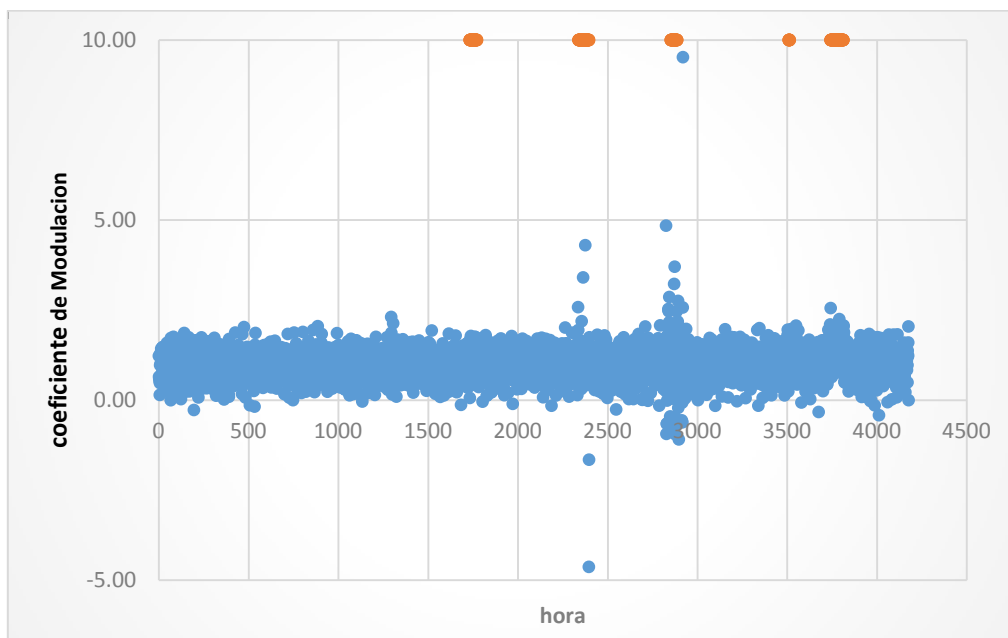


Figura 34. Coeficiente de Modulación Sector 1 dataset 2.

Como podemos observar, los puntos de color marrón indican los instantes donde están localizados los ataques, que evidentemente coinciden con aquellos puntos que no se ajustan a la tendencia del comportamiento del consumo, por lo cual el modelo está detectando anomalías en varios tramos del proceso de una manera muy efectiva.

Con respecto a lo antes mencionado, podemos comparar este comportamiento con aquel que se presentó en el dataset 1, el cual muestra una disposición similar a excepción de aquellos puntos que no hacen parte de la tendencia. La grafica del Sector 1 para el dataset 1 se muestra en la **Figura 35**.

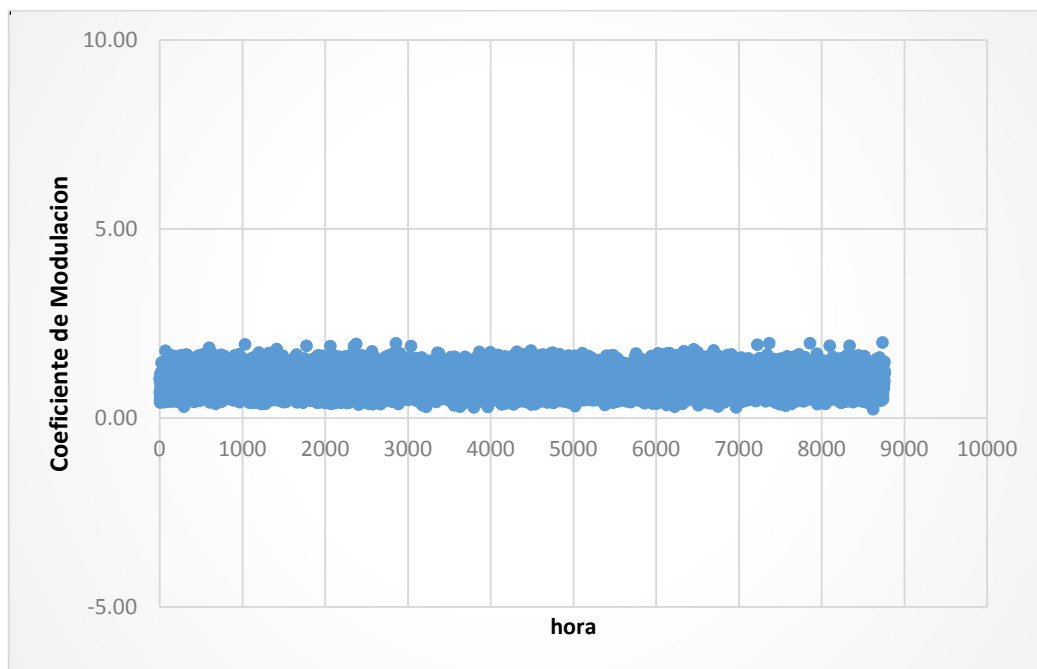


Figura 35. Coeficiente de Modulación normal Sector 1.

Para este sector, los valores del dataset 1 oscilan entre un máximo de 2 y un mínimo de 0.23, por lo cual, atendiendo al criterio de evaluación de los patrones de consumo, todos aquellos fuera de este rango harán parte de posibles anomalías en el sistema.

4.1.2 VALIDACION DE CONSUMOS SECTOR 2

En cuanto al Sector 2, al que pertenecen las bombas PU4, PU5 y el tanque T3, después de introducir los datos en el modelo y realizar la posterior simulación de los consumos, se encontró la siguiente curva de modulación:

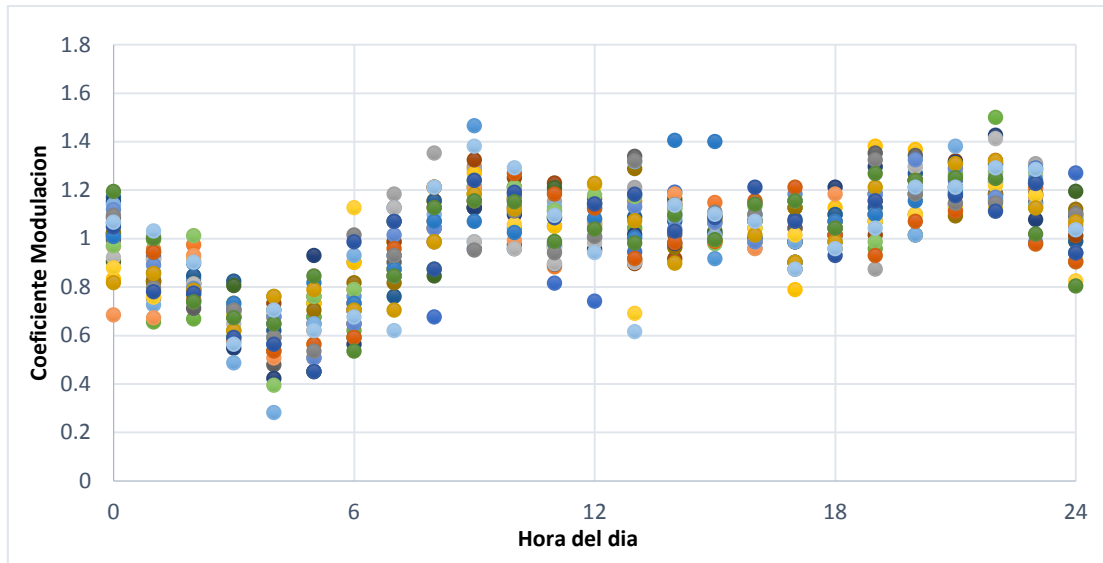


Figura 36. Curva de Modulación Sector 2 dataset 2.

Observando en principio la **Figura 36**, este sector muestra un mejor comportamiento a la tendencia de la curva, siendo muy semejante a la **Figura 6**. Así mismo la **Figura 37** muestra el análisis completo, donde se ve una disposición relativamente normal de los datos durante todo el tiempo del análisis.

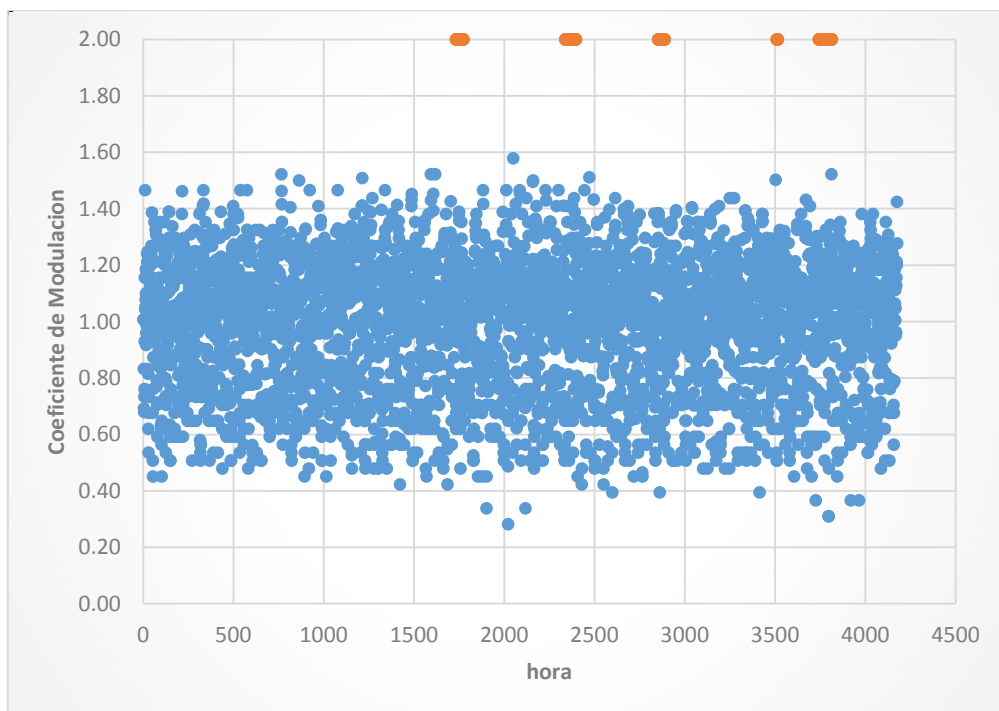


Figura 37. Coeficiente de modulación Sector 2 dataset 2.

Siguiendo con la comprobación, hay que observar la disposición de los registros a partir del dataset 1, de los cuales igual que en el sector anterior, se calculará los máximos y mínimos con el propósito de comparar estos dos comportamientos. A continuación, se muestra la tendencia correcta para el Sector 2:

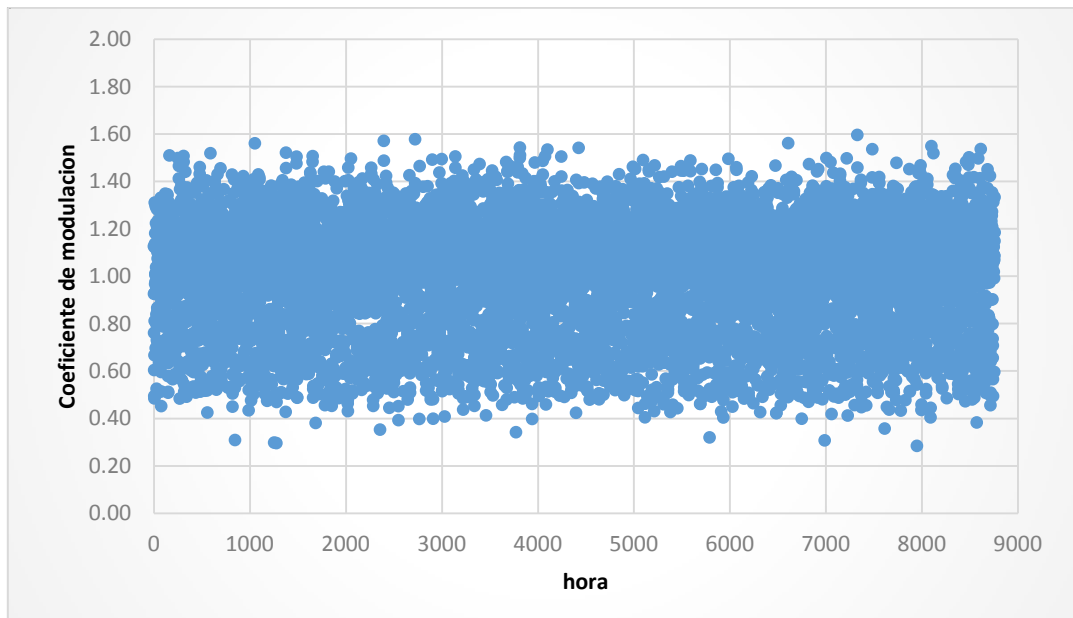


Figura 38. Coeficiente de Modulación normal Sector 2.

La **Figura 38** muestra unos rangos de operatividad similares y a través del criterio de evaluación, se ha determinado el rango de funcionamiento normal el cual tiene un valor mínimo de 0.23 y un valor máximo de 1.99. De esta manera, al parecer existirán pocos o ningún valor que pueda calificarse como sospechosos debido a que las dos tendencias son muy parecidas.

4.1.3 VALIDACION DE CONSUMOS SECTOR 3

Ingresando los datos correspondientes al modelo de detección, se ha logrado obtener la curva de modulación para el Sector 3, donde se evidencia también datos muy alejados a la tendencia de curva, datos que posiblemente sean anomalías como se aprecian en la **Figura 39**.

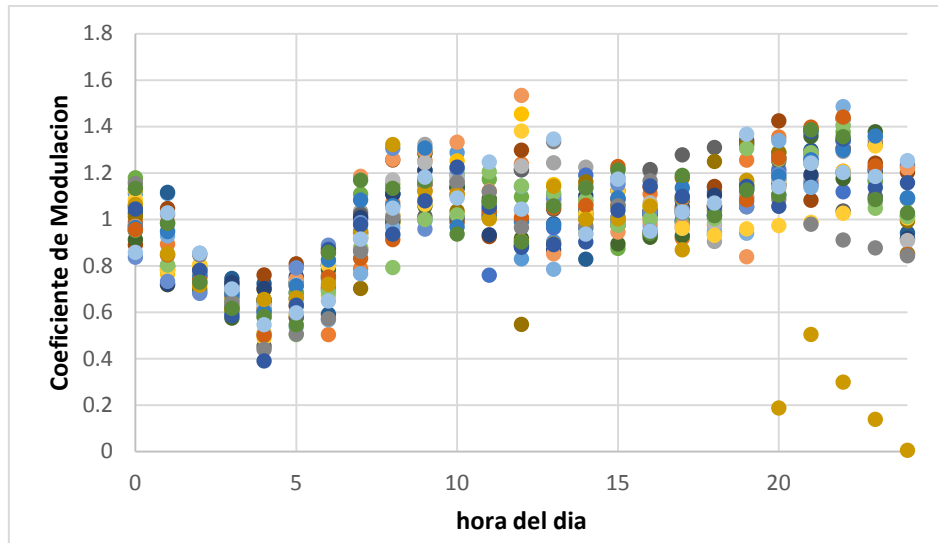


Figura 39. Curva de Modulación Sector 3 dataset 2.

Así mismo, observando la **Figura 40**, el modelo ha detectado instantes donde los valores no corresponden a un comportamiento normal, por ejemplo, entre la hora 3500 y 4000 de la simulación, claramente concuerdan dos intervalos de datos sospechosos, con los instantes donde aparecen las anomalías en la red marcadas con color marrón.

En el mismo sentido, la **Figura 41** muestra la tendencia del dataset 1, el cual concuerda relativamente con la mayoría de valores del segundo conjunto de datos, demostrando así una alta probabilidad que aquellos valores que sean menores al valor mínimo de 0.11 y mayores al valor máximo de 1.58 sean fallos en el sistema.

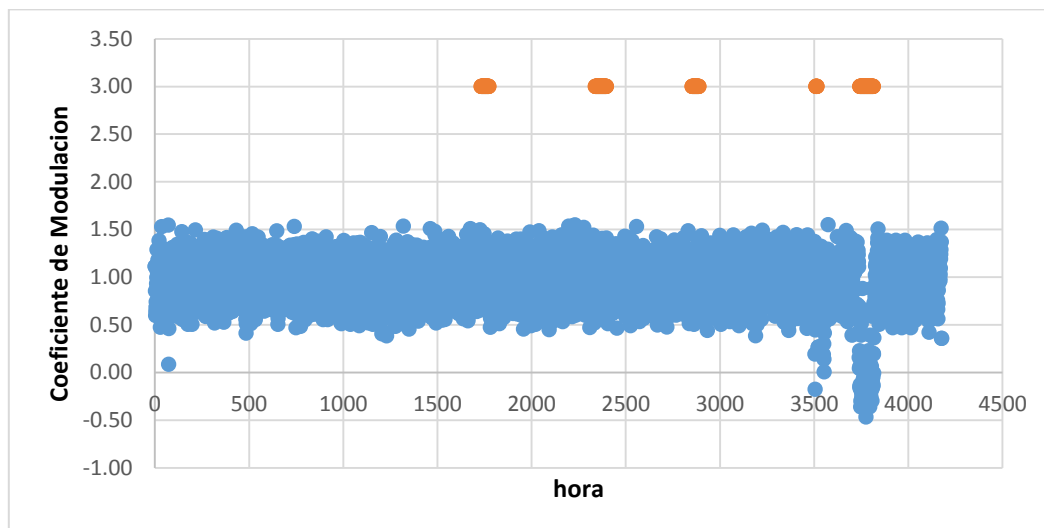


Figura 40. Coeficiente de Modulación Sector 3 con anomalías.

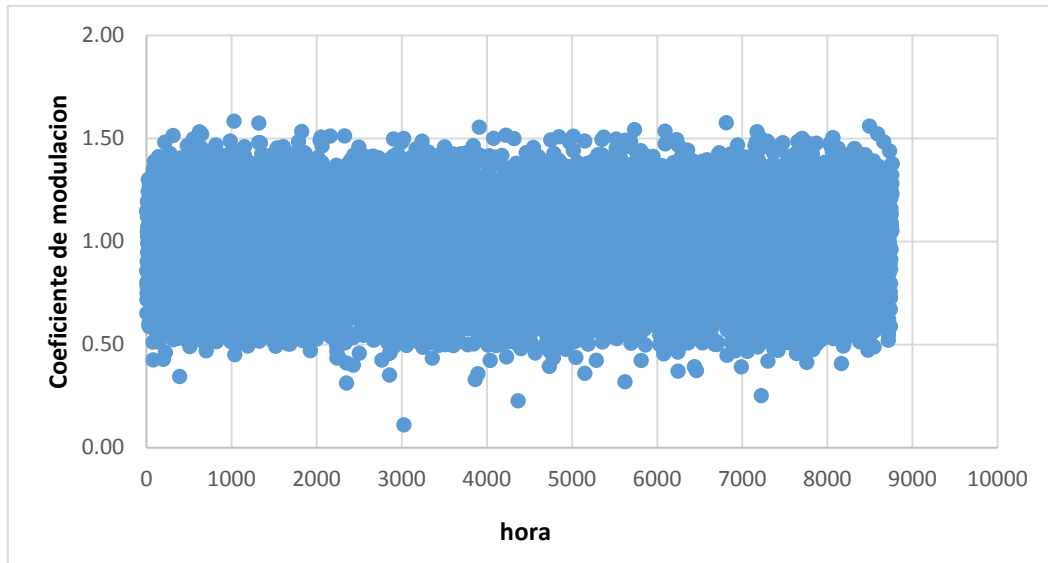


Figura 41. Coeficiente de Modulación normal Sector 3.

4.1.4 VALIDACION DE CONSUMOS SECTOR 4

Este sector al igual que el Sector 2, presenta un comportamiento relativamente normal como se observa en la **Figura 42**, donde al parecer no existe gran variabilidad entorno a la tendencia de la curva. También la **Figura 43** nos muestra el mismo comportamiento y al parecer no existirán valores que coincidan con los periodos de anomalías marcados en color marrón, no obstante, todos aquellos valores deben validarse de acuerdo al criterio establecido, lo cual implica establecer un rango correcto de operación y de esa forma seleccionar los datos sospechosos.

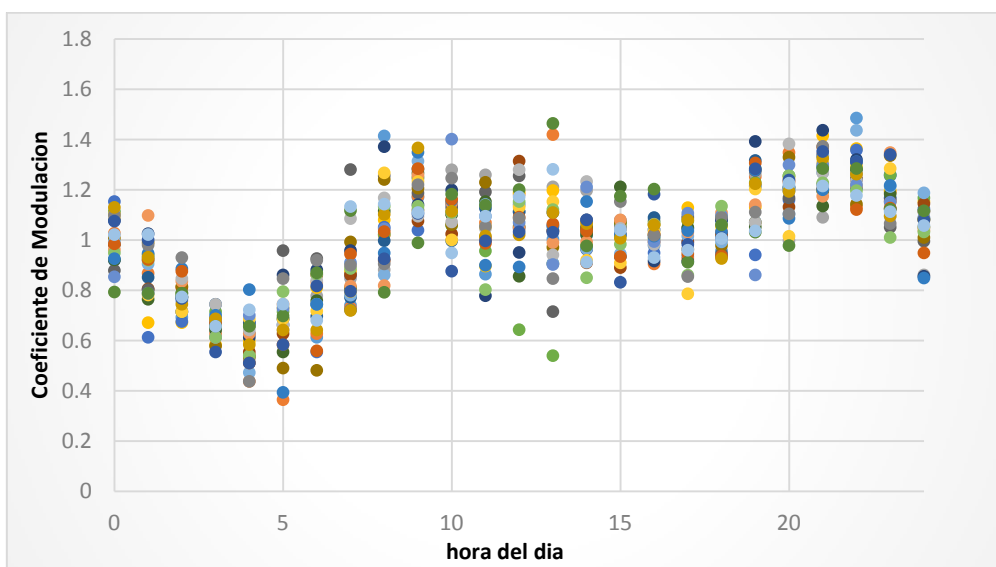


Figura 42. Curva de Modulación del Sector 4 dataset 2.

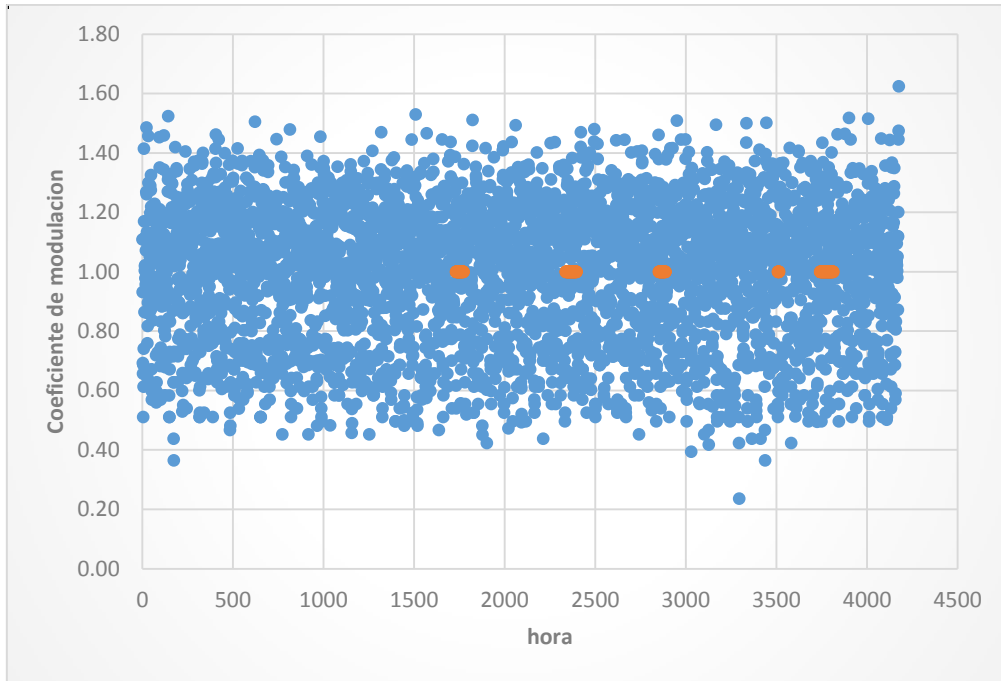


Figura 43. Coeficiente de Modulación Sector 4 dataset 2.

El rango encontrado en el dataset 1 para este sector, como se muestra en la **Figura 44**, tiene un valor mínimo de 0.065 y un máximo de 1.62, por lo cual, todos aquellos valores que no estén dentro de este rango, harán parte de la lista de sospechosos para este sector.

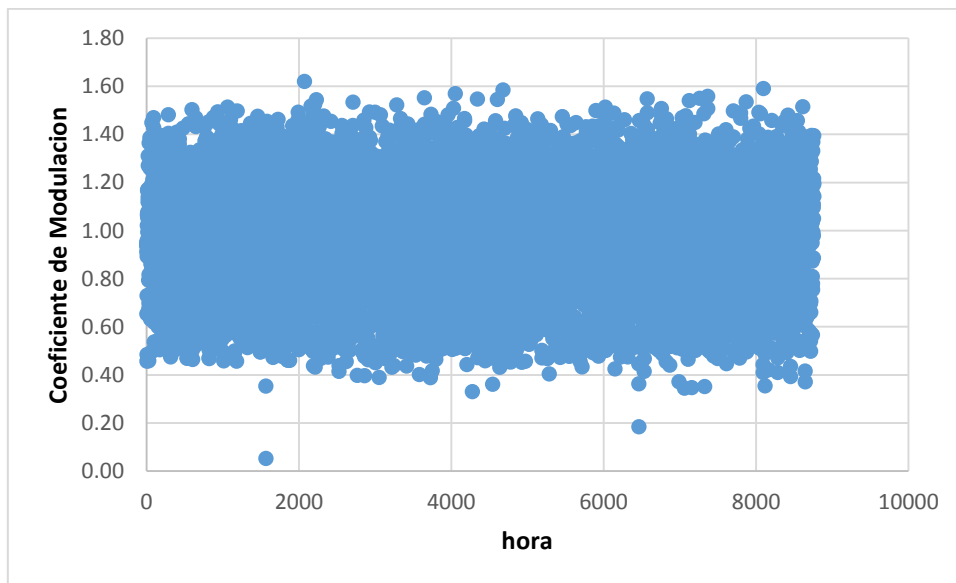


Figura 44. Coeficiente de Modulación normal Sector 4

4.1.5 VALIDACION DE CONSUMOS SECTOR 5

Después de ingresar los datos al modelo se encontró en el Sector 5 un comportamiento relativamente normal cuando se representó la curva de modulación como se muestra en la **Figura 45**, con rangos evidentemente más estrechos que en los demás sectores estudiados.

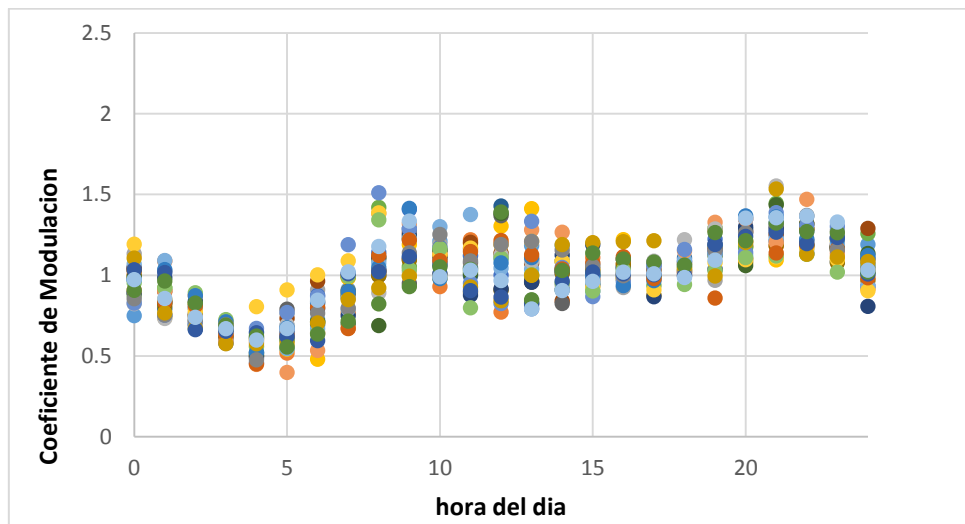


Figura 45. Curva de Modulación Sector 5 dataset 2

No obstante, al observar el comportamiento del patrón de consumo durante toda la simulación, se pudo apreciar que había valores que se desajustaban de la tendencia general del Sector, como se observa en la **Figura 46**, donde evidentemente aquellos valores que están fuera de la tendencia general coinciden con las horas de anomalías presentadas en el sistema.

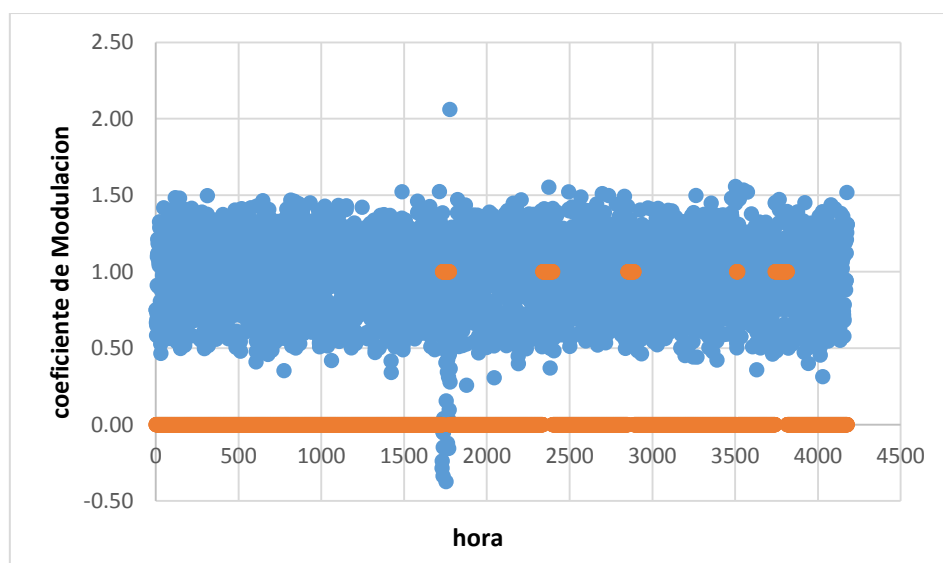


Figura 46. Coeficiente de Modulación Sector 5 dataset 2.

Para poder determinar por lo tanto los valores sospechosos, igual que en los demás Sectores hay que determinar el rango de funcionamiento a partir del dataset 1 y de esta manera seleccionar los datos que no hagan parte del mismo. A continuación, se presenta la simulación para el primer conjunto de datos:

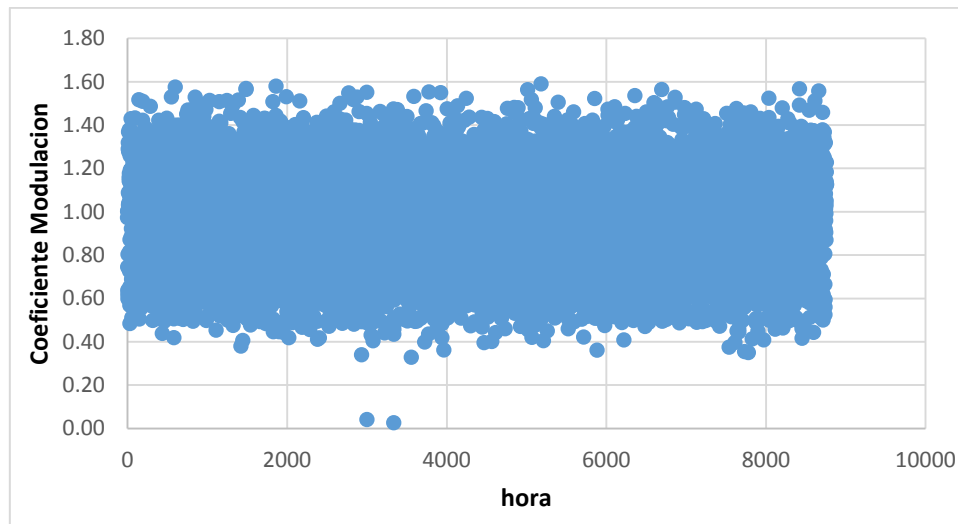


Figura 47. Coeficiente de Modulación normal Sector 5

A partir de esto podemos determinar el rango normal de funcionamiento para el patrón de consumo, el cual tiene un valor mínimo de 0.48 y un máximo de 1.55. Conocido estos valores, todos aquellos consumos que estén fuera de los límites se etiquetaran como posibles anomalías.

4.2 VALIDACION DE LAS ESTACIONES DE BOMBEO.

De igual manera que con los consumos, las estaciones de bombeo son un punto crítico en todo el sistema, puesto que todo el comportamiento de la red está directamente relacionado al caudal que cada estación bombea, es por eso que para evaluar el comportamiento de las bombas se ha creado un criterio que involucra una comparativa de las curvas características para los dos conjuntos de datos. En este sentido, el criterio es el siguiente:

- Todos aquellos datos que estén fuera de los rangos de funcionamiento impuestos por las curvas características de las bombas en condiciones normales, se clasificarán como sospechosos.

4.2.1 VALIDACION DE LA ESTACION DE BOMBEO SECTOR 1

Este sector cuenta con tres bombas, de las cuales dos están en funcionamiento durante todo el intervalo de tiempo registrado por los datos de telemetría, además la estación de bombeo del

Sector 1 abastece a los demás sectores y según los datos calculados en el dataset 1 y observados en la **Figura 17** y **Figura 18**, los rangos de funcionamiento son:

BOMBA	RANGO1(l/s)	RANGO2(l/s)
PU1	85.49-100.12	109.98-122.70
PU2	85.51-100.14	-

Tabla 13. Rango de funcionamiento de las bombas Sector1.

Conocido los rangos normales de operación de las bombas, es necesario conocer que curva característica resulto en el dataset 2 y con esto, contrastar los resultados obtenidos al principio con los que ahora nos arroja el modelo. De esta forma las curvas características en el dataset 2 para las bombas PU1 y PU2 son las siguientes:

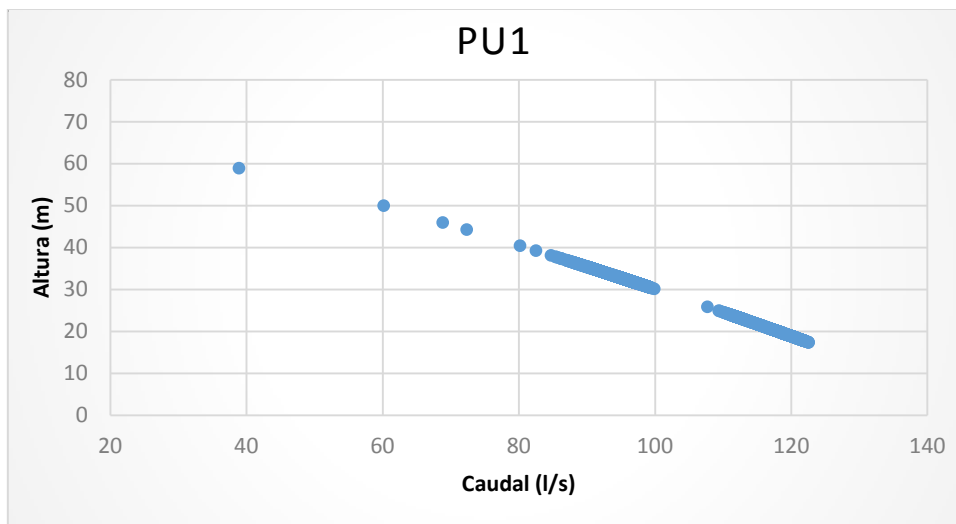


Figura 48. Curva característica de la bomba 1 con anomalías.

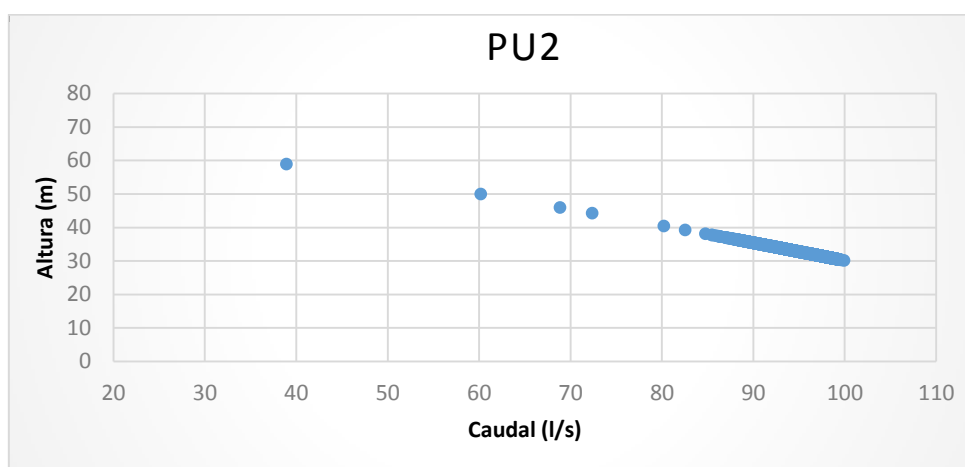


Figura 49. Curva característica de la bomba 2 con anomalías.

Como se observa en las figuras anteriores, el modelo nos está arrojando varios datos que no se ajustan al rango de funcionamiento normal de las bombas, por ejemplo, para la bomba PU1 y PU2 es raro encontrar caudales muy bajos de funcionamiento, el mínimo para las dos está entorno a los 39 l/s lo cual genera sospechas sobre su comportamiento. Conocido entonces la curva de la bomba y con el criterio de evaluación establecido, todos aquellos valores que no estén dentro del rango propuesto en la Tabla 2, se etiquetaran como sospechosos.

4.2.2 VALIDACION DE LA ESTACION DE BOMBEO SECTOR 2

En cuanto al Sector 2, aunque están asignadas dos bombas solo una de ellas está en funcionamiento y su curva característica se ve representada en la figura 18, donde se muestra claramente un rango de funcionamiento relativamente normal con caudales entorno a las 30 y 40 l/s. De esta manera para crear el punto de comparación para la posterior evaluación del criterio, se determinó los rangos exactos de funcionamiento para este sector como se indica en la **Tabla 14**.

BOMBA	RANGO 1 (l/s)	RANGO 2 (l/s)
PU4	30.47-38.29	-
PU5	-	-

Tabla 14. Rango de funcionamiento de las bombas Sector 2

Si siguiendo con la evaluación del criterio, es necesario conocer el comportamiento de la bomba PU4 en el dataset 2. En este sentido, la curva característica para esta bomba con datos anómalos se muestra en la **Figura 50**.

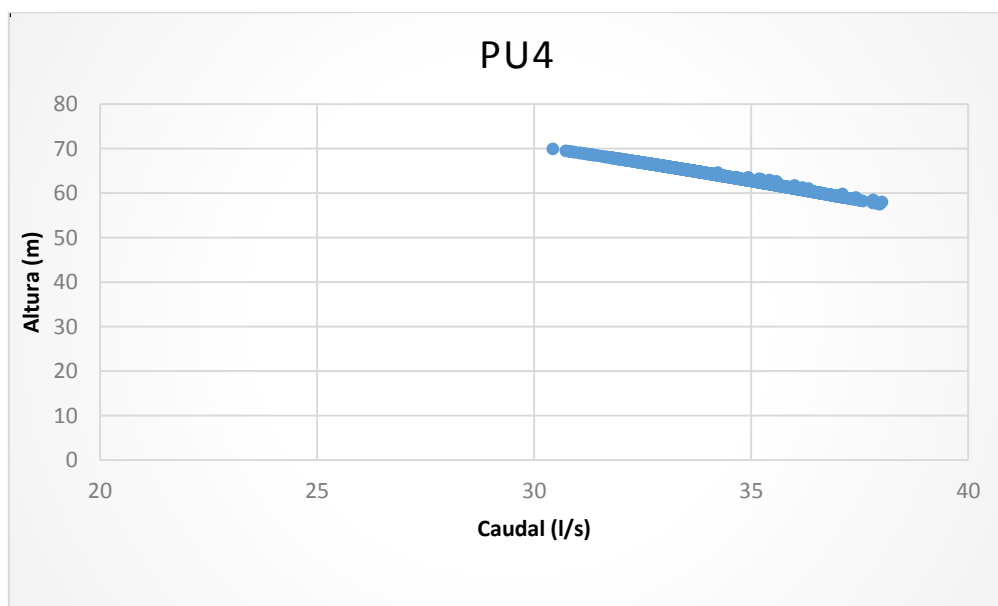


Figura 50. Curva característica de la bomba 4 dataset 2.

La figura anterior nos muestra un comportamiento relativamente normal en la bomba PU4, con caudal mínimo de 30.43 l/s y máximo de 38.01 l/s, por lo cual existirán muy pocos valores sospechosos para este Sector en comparación a los valores registrados en el primer conjunto de datos y mostrados en la **Tabla 14**.

4.2.3 VALIDACION DE LA ESTACION DE BOMBEO SECTOR 3

El sector 3 está conformado por dos bombas, las cuales están en funcionamiento en paralelo durante algunos periodos de tiempo como se miró en la Figura 19 y 20, donde existen dos rangos de funcionamiento para la bomba PU6 y un solo rango para la bomba PU7. La **Tabla 15** muestra estos rangos de operación normal para esta estación de bombeo.

BOMBA	RANGO 1 (l/s)	RANGO 2 (l/s)
PU6	35.75 – 37.80	-
PU7	35.73 – 37.78	45.1 – 51.65

Tabla 15. Rango de funcionamiento de las bombas Sector 2

Con el rango de funcionamiento claramente establecido, como en los sectores anteriores, se ha de observar el comportamiento de la estación de bombeo 3 y de esta forma determinar datos sospechosos que no estén dentro del rango normal de operación de las bombas. La **Figura 51** y **Figura 52** muestra las curvas características para las bombas 6 y 7.

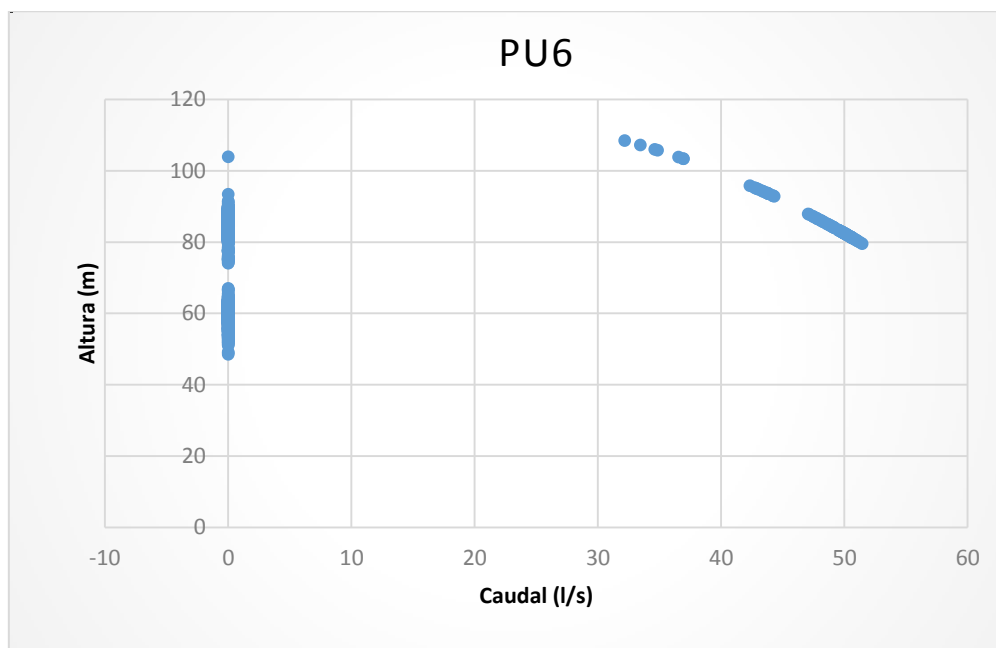


Figura 51. Curva característica bomba 6 con anomalías.

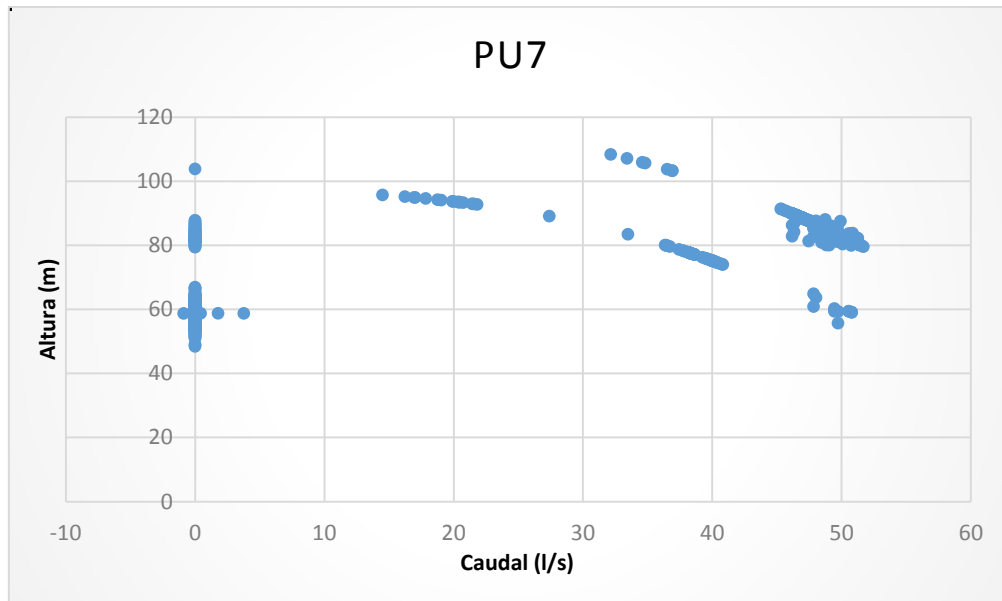


Figura 52. Curva característica de la bomba 7 con anomalías.

Observando las figuras anteriores, podemos a simple vista determinar que algo sospechoso está pasando en esta estación de bombeo. Primeramente, la bomba 6 tiene muchos rangos de operación que, aunque están dentro de la curva de la bomba, no corresponden a un funcionamiento normal, debido a que gran parte de los valores sobrepasan el valor máximo de 37.80 l/s.

En cuanto a la bomba 7, los valores muestran aún más un comportamiento anormal, donde se generan tres curvas características de las cuales la primera corresponde a valores menores a 5 l/s, que en la realidad son datos casi imposibles de generar, puesto que existen caudales negativos y a la vez muy pequeños para que una bomba funcione en condiciones óptimas. La segunda curva característica funciona a alturas más elevadas, sin embargo, esta no corresponde a la curva mostrada en la **Figura 21** y pareciera que estuviera funcionando a una velocidad de giro diferente. Para saber exactamente lo que sucede en esta bomba se ha de analizar si por alguna razón la bomba pudo estar funcionando con el uso de un variador de frecuencia a una velocidad distinta a la nominal.

Para determinar la velocidad de funcionamiento de la bomba se calculó la ecuación de la misma a partir de la herramienta línea de tendencia usada en el programa Excel para el dataset 1 en la bomba 7 y de esta manera encontrar el porcentaje de reducción de velocidad a la cual podría estar sometida la bomba. Después de algunas iteraciones el porcentaje que más se ajusta al comportamiento de esta segunda curva, es el 90% y genera una curva muy similar con respecto a la presentada en este último conjunto de datos. Las curvas se muestran a continuación:

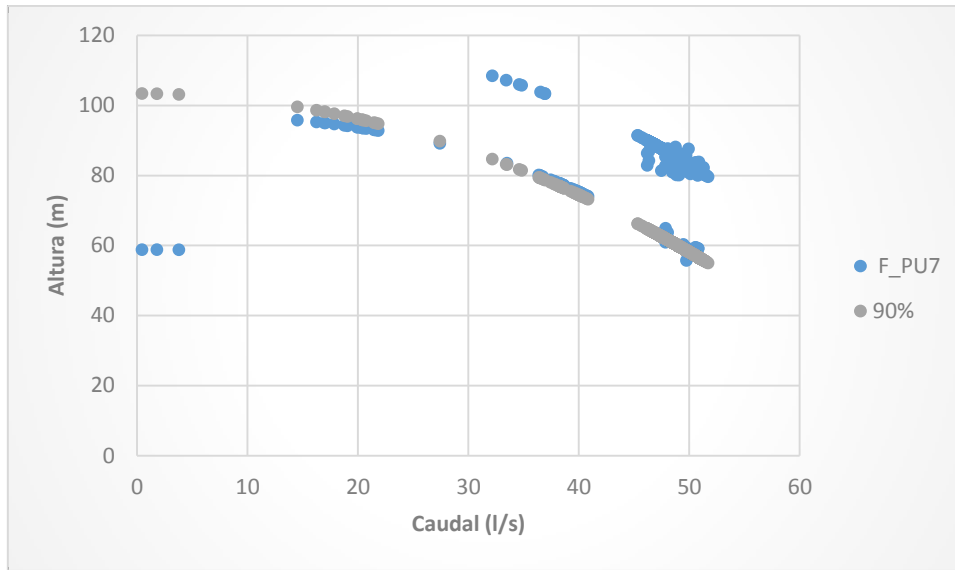


Figura 53. Variación de la velocidad de giro de la bomba 7.

Según la figura anterior, la mejor explicación a este comportamiento es que la bomba está siendo manipulada a través de su velocidad de giro, por lo cual funcionando al 90% de su velocidad la bomba en particular bombea menos caudal que en condiciones normales. También es importante mencionar que BATADAL asume cada una de las bombas como bombas de velocidad fija, por lo que este funcionamiento contradice aquella afirmación; en conclusión, se asumirá este comportamiento como dudoso, por lo cual aquellos valores que no hagan parte de la curva característica nominal serán separados como sospechosos.

Por último, está la curva característica superior, que evidentemente funciona al 100 % de su velocidad nominal, por lo que en estos datos si aplicaremos el criterio establecido, etiquetando como sospechosos aquellos valores que no hagan parte del rango de funcionamiento y que no estén dentro de la curva característica de la bomba.

4.2.4 VALIDACION DE LA ESTACION DE BOMBEO SECTOR 4

Este Sector posee una estación de bombeo con dos bombas de las cuales como en el sector 2 tan solo una de ellas está en funcionamiento. Por lo cual se ha de identificar los rangos de funcionamiento normales para poder así evaluar el criterio. La **Tabla 16** reúne esta información.

BOMBA	RANGO 1 (l/s)	RANGO 2 (l/s)
PU8	32.24 – 39.13	-
PU9	-	-

Tabla 16. Rango de funcionamiento de las bombas Sector 2

Ahora será necesario contrastar este rango con la curva encontrada en el dataset 2. La Figura 45 nos muestra la curva característica de la bomba PU8, donde se observa evidentemente varios valores muy lejanos del rango de funcionamiento normal, por lo cual todos estos que estén por fuera de los valores máximos y mínimos de la **Tabla 16** serán marcados como sospechosos.

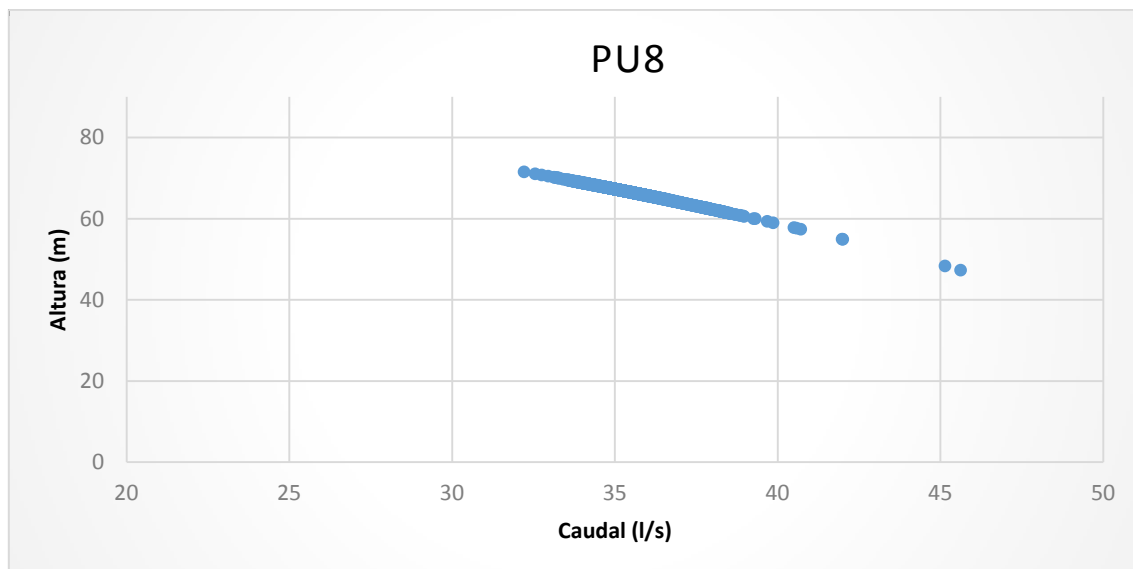


Figura 54. Curva característica de la bomba 8 con anomalías.

4.2.5 VALIDACION DE LA ESTACION DE BOMBEO SECTOR 5

Al sector 5 lo conforma una estación de bombeo con las bombas PU10 y PU11; las cuales como se miró en la **Figura 23** y **Figura 24** comparten un rango de funcionamiento en paralelo en muy pocos valores, es decir, la bomba PU10 está encendida en escasos momentos. Los rangos de funcionamiento normal para las dos bombas se fijan en la Tabla 17.

BOMBA	RANGO 1 (l/s)	RANGO 2 (l/s)
PU8	32.24 – 39.13	-
PU9	-	-

Tabla 17. Rango de funcionamiento de las bombas Sector 2

El segundo dataset 2 por otro lado nos muestra comportamientos muy particulares y a la vez sospechosos como se observa en las **Figura 55**, donde en el caso particular de la bomba 10, se ve claramente valores que no se ajustan a la tendencia de la curva característica, observando valores que están por encima y debajo de la misma y que no tienen un comportamiento relacionado a ningún tipo de maniobra hidráulica en bombas. Por otro lado, la bomba 11 ha funcionado muchas

más horas en comparación al dataset 1, sin embargo, coincide perfectamente a la curva de la bomba 10, demostrando así un funcionamiento en paralelo de ambas.

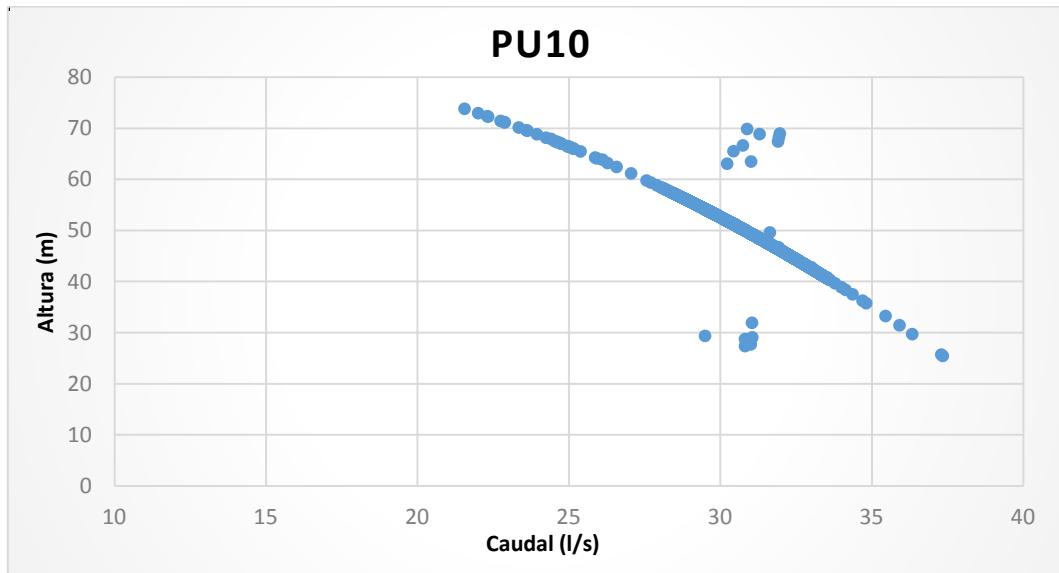


Figura 55. Curva característica de la bomba 10 con anomalías.

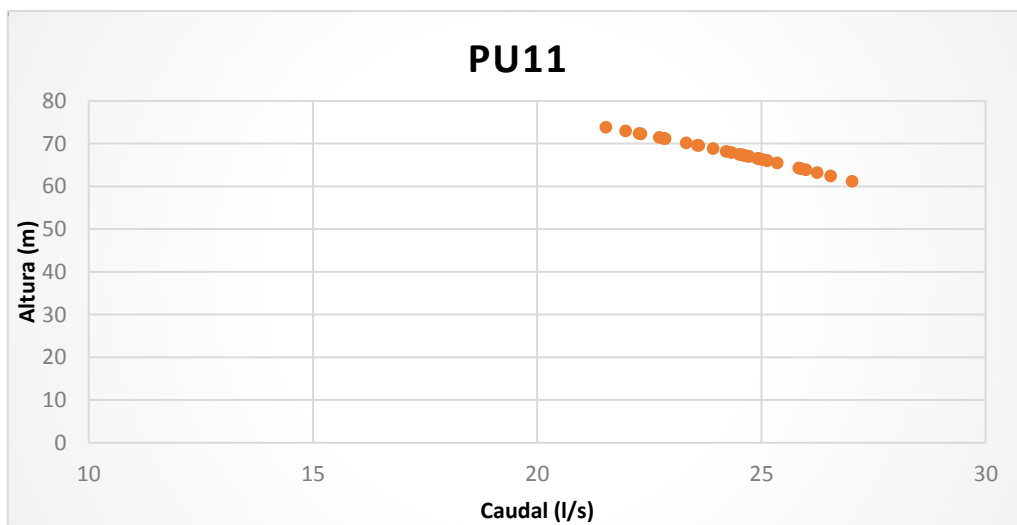


Figura 56. Curva Característica de la bomba 11 con anomalías.

En este mismo sentido, se ha de evaluar por lo tanto el criterio establecido para las dos bombas y particularmente en la bomba 10, aunque los valores estén dentro del rango, mucho de ellos no se consideraran datos correctos puesto que no corresponden a la curva de la Bomba en condiciones de funcionamiento normal.

4.3 VALIDACION DE LOS TANQUES

Para la validación de datos de los tanques en la red C-Town, se introdujo los valores del dataset 2, de forma que la detección de anomalías estará basada en la comparación de los nuevos valores con los límites superiores e inferiores encontrados en el dataset 1. En consecuencia, el criterio de evaluación será el siguiente:

- Todos aquellos Niveles que estén fuera de los límites inferior y superior encontrados en el dataset 1 serán etiquetados como sospechosos.

Es decir que cada Sector tendrá un valor diferente de evaluación para cada Tanque y se espera encontrar una mayor cantidad de anomalías debido a que son puntos muy críticos en toda la red. Empezando con el análisis, el Tanque 1 presenta el siguiente comportamiento:

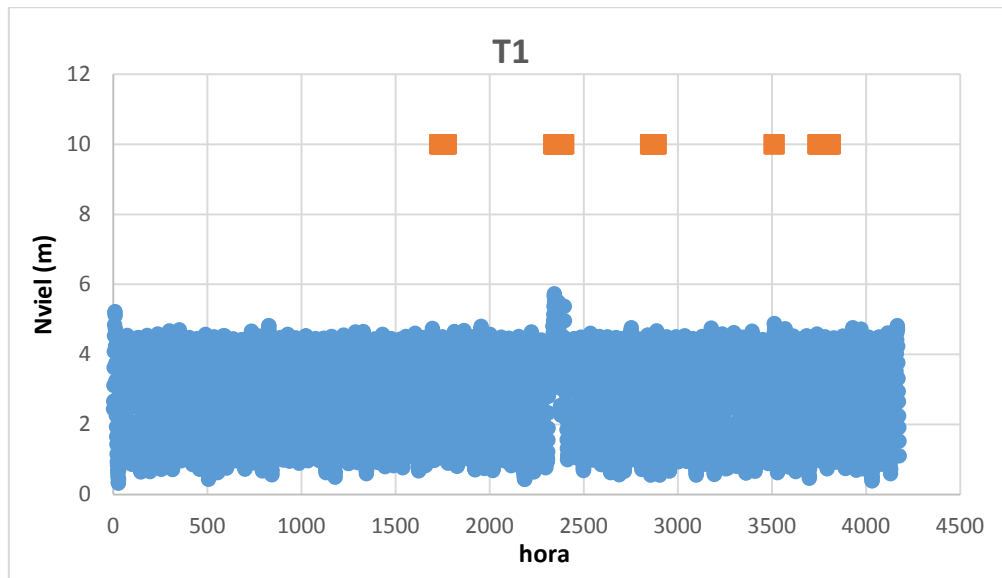


Figura 57. Nivel del Tanque 1 dataset 2

Para el Tanque 1 según la Figura 57, se ve perfectamente valores anómalos entre la hora 2000 y 2500, donde es notorio que estos sobrepasan el valor límite superior de la Tabla 8 y concuerdan perfectamente con uno de los periodos de fallo del Sistema marcados en color marrón, por lo que todos aquellos valores que no estén dentro de los rangos establecidos a partir del dataset 1 serán marcados como sospechosos.

Siguiendo con el Sector 1, la **Figura 58** muestra el funcionamiento del Tanque 2, donde los valores están más ajustados al comportamiento normal de la red, por lo cual se esperaran pocos o ningún valor sospechoso para este Tanque.

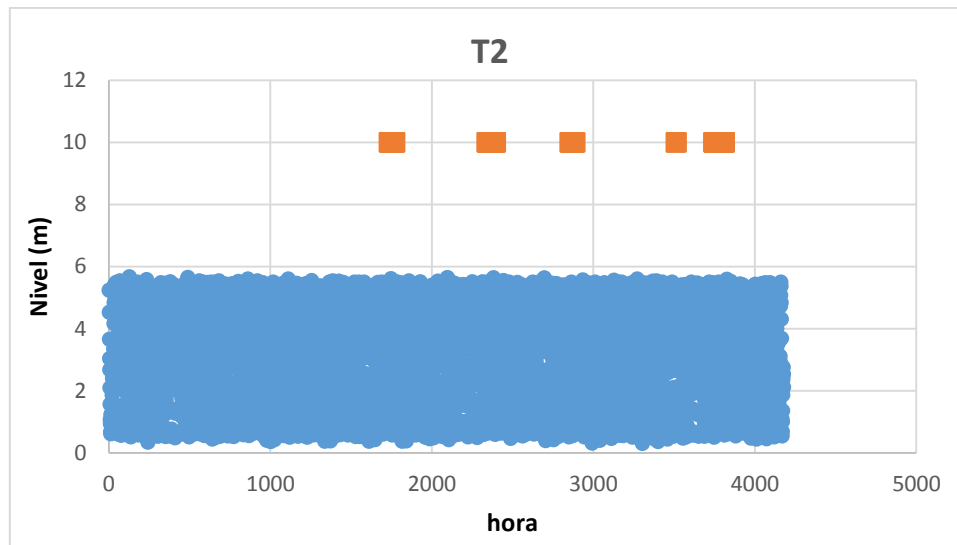


Figura 58. Nivel del tanque 2 dataset 2

Siguiendo con la validación, el Sector 2 contiene al Tanque 3, el cual cómo podemos observar en la **Figura 59** posee una tendencia muy similar a la encontrada en la **Figura 28**, por lo que igual que en el Tanque 2, brindara muy poca información para la detección de algún fallo en la red.

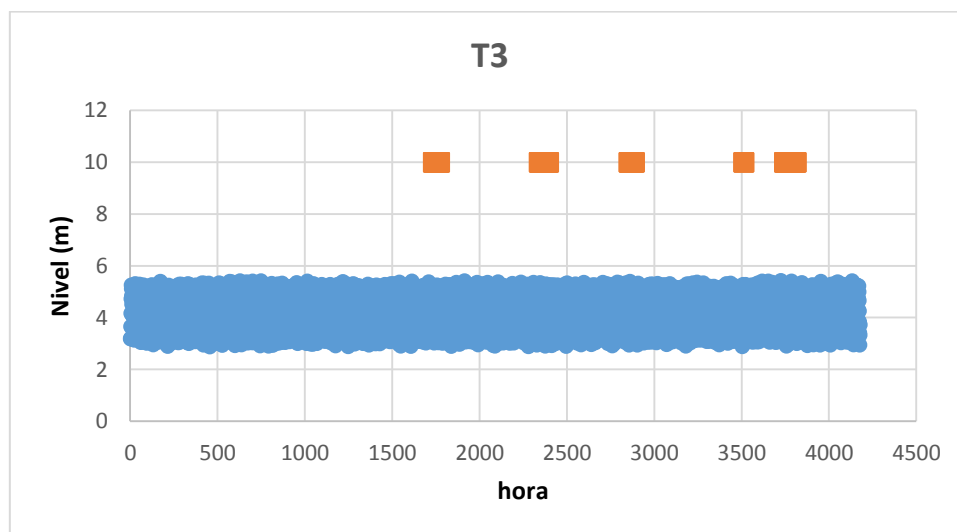


Figura 59. Nivel del Tanque 3 dataset 2

Para el Sector 3, al que pertenece el Tanque 4, vemos a través de la **Figura 60** una disposición de los datos similar a la del dataset 1 representada en la **Figura 29**, encontrando igualmente variabilidad en los datos aunque en este caso si es un tanto visible valores que se alejan más de lo habitual precisamente en los dos últimos instantes de fallo. De esta manera y aplicando el criterio de evaluación, todos aquellos que no hagan parte del rango de Niveles serán marcados como sospechosos.

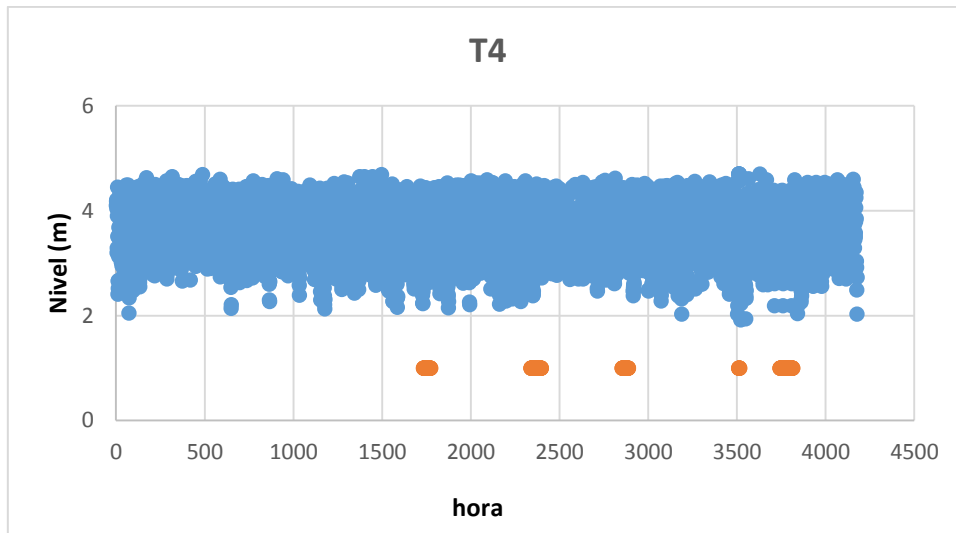


Figura 60. Nivel del Tanque 4 dataset 2

Así mismo para el Sector 4, el Tanque 5 tiene igualmente un comportamiento relativamente normal en comparación al funcionamiento encontrado a partir del dataset 1. En este caso, según los datos de la **Figura 61** tampoco se obtendrá mucha información que nos lleve a detectar las anomalías.

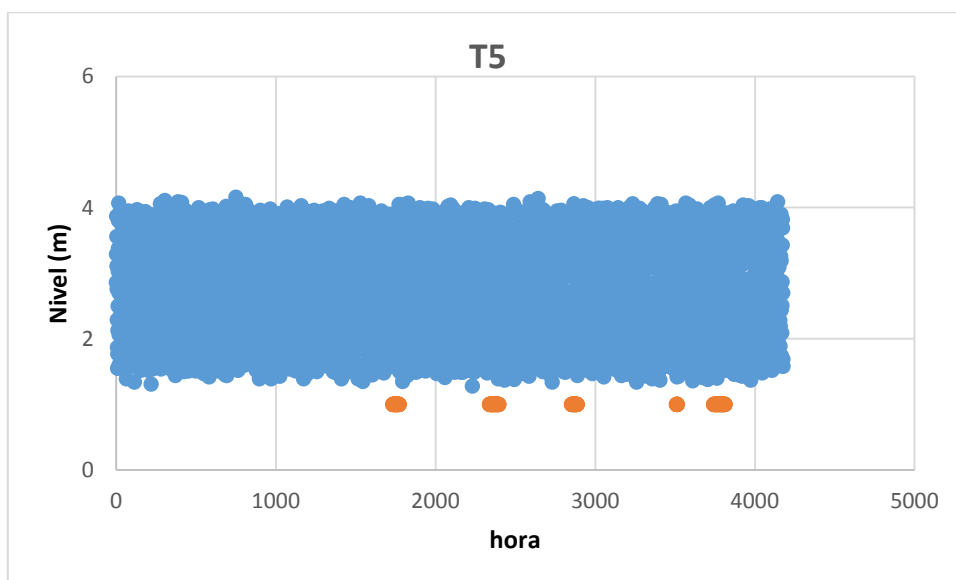


Figura 61. Nivel del Tanque 5 dataset 2

Por último, para el Sector 5, al cual pertenece el Tanque 6 y 7, la tendencia del primero no muestra comportamientos sospechosos, por lo tanto, serán pocos o ningún valor que pueda aportar a la detección de alguna anomalía.

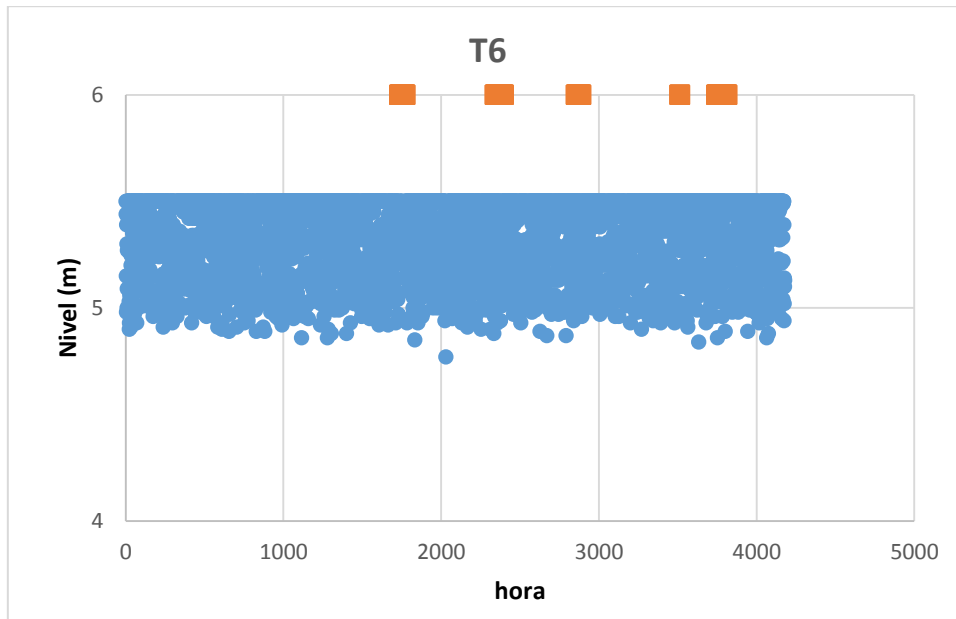


Figura 62. Nivel del Tanque 6 dataset 2

Por otro lado en la **Figura 63**, el Tanque 7 muestra la misma variabilidad respecto al dataset 1 de la **Figura 32**; sin embargo los valores son aún más inferiores y atendiendo al criterio de validación establecido al comienzo de este capítulo, todos aquellos valores que están fuera de los límites superior e inferior serán considerados como posibles fallos del sistema.

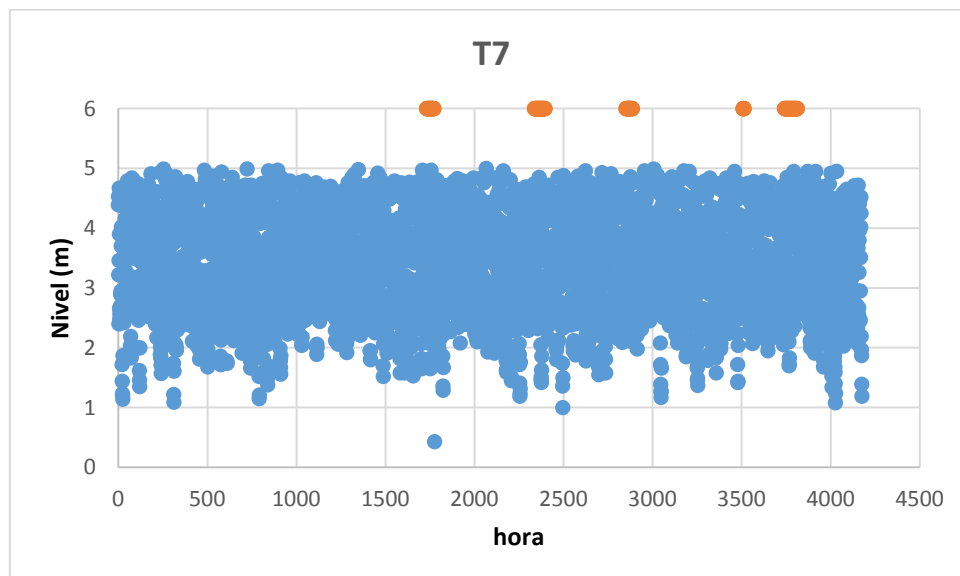


Figura 63. Nivel del Tanque 7 dataset 2



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA

máster en ingeniería
hidráulica y medio ambiente
mihma

5. RESULTADOS

El objetivo del proyecto como se habló al inicio, es poder establecer criterios de detección de anomalías accidentales o provocadas, de tal manera que la red pueda responder lo más pronto posible a estos tipos de fallos. En este contexto, este modelo de detección ha de recoger todas aquellas detecciones producidas por la evaluación de los criterios en cada una de las medidas usadas y bastara que se alerte una sola vez para considerar un estado inseguro en el sistema. Del mismo modo es importante mencionar que de acuerdo a los valores de evaluación de la Matriz de resultados, se dará prioridad a la detección de todas las anomalías por encima de la disminución de falsos positivos.

Para la recolección de los resultados, se analizarán aquellos valores de acuerdo al criterio de evaluación, empezando por los patrones de consumos y finalizando con los tanques, de tal forma que al final cada hora tendrá un valor de numero de sospechas, las cuales serán comparadas con los instantes donde la red realmente se encuentra en ataque. Por otro lado, se evaluarán los criterios establecidos de acuerdo a su efectividad para detectar anomalías y el grado de importancia que tiene sobre los demás. Por último, se obtendrán los resultados de la Matriz de confusión que determinará la efectividad final del Modelo.

5.1 ANOMALIAS EN LOS CONSUMOS.

Para la identificación de los valores sospechosos para este criterio, se planteó inicialmente un balance de caudal, el cual en algunos sectores cambiaba de acuerdo al comportamiento de los arranques y paradas de las bombas. Con esto sucesivamente se calculaba los consumos y dividiendo por el valor medio de toda la serie encontrada, se calculaba el patrón de consumo respectivo. Obtenido esto, se obtenía la curva de modulación y los rangos correctos de consumo como base para la detección de posibles anomalías. Posteriormente se introdujo al modelo de detección los valores del dataset 2, con las anomalías que el modelo debía detectar a partir de la evaluación de los criterios. En esta parte se ha de recoger los resultados del modelo en cuanto a los consumos, clasificando la cantidad de alertas que arroja el modelo, como los aciertos que este tuvo durante la simulación. Esta información la podemos observar en la **Tabla 18**.

DESCRIPCION	PATRONES DE CONSUMO				
	1	2	3	4	5
SOSPECHOSOS	52	0	61	0	36
ACIERTOS	21	-	58	-	19
EFFECTIVIDAD	40.3%	-	95%	-	52.7%

Tabla 18. Resultados de la detección por Consumos

Observando la tabla anterior, se puede afirmar que en general la eficiencia del criterio basado en los patrones de consumo es muy variable, por ejemplo, en el caso del patrón del Sector 1, que es uno de los que más presenta valores sospechosos, se detectaron un poco más del 40% de las anomalías reales en base al número de sospechas presentadas. Algo parecido ha sucedido con los resultados del Sector 5, sin embargo, los que no presentaron detecciones fueron los Patrones 2 y 4, los cuales de antemano se preveía su detección nula debido a la buena tendencia que generaron las variables analizadas. De esta manera el patrón de consumo más eficiente es el del Sector 3, con una efectividad del 95 %.

Por otro lado, también se puede distinguir no solo la efectividad de los patrones, sino también la localización de cada uno dentro del contexto general de las anomalías, por ejemplo en la **Figura 64** se puede observar que el Patrón 1 ha detectado valores anómalos del segundo, tercero y quinto periodo de ataque, sin embargo, este no ha detectado el primero y el tercer periodo de ataques. Estos resultados son evidentes cuando se observó en la **Figura 34** el comportamiento del patrón de consumo durante toda la simulación, el cual mostraba tres grupos de valores que se alejaban de la tendencia general, en consecuencia, el criterio ha detectado en tres grupos de fallos.



Figura 64. Localización de Aciertos Patrón 1.

Así mismo el Patrón 3, con el porcentaje de efectividad mayor, ha detectado casi todos los valores del periodo 5 de fallos, no obstante, los demás periodos de anomalías fueron desconocidos para este patrón y coincide precisamente con lo observado en la **Figura 40**, donde se observaba claramente los valores anómalos entre la hora 3500 y 4000. La Figura 65 muestra el comportamiento de las detecciones para el Patrón 3.

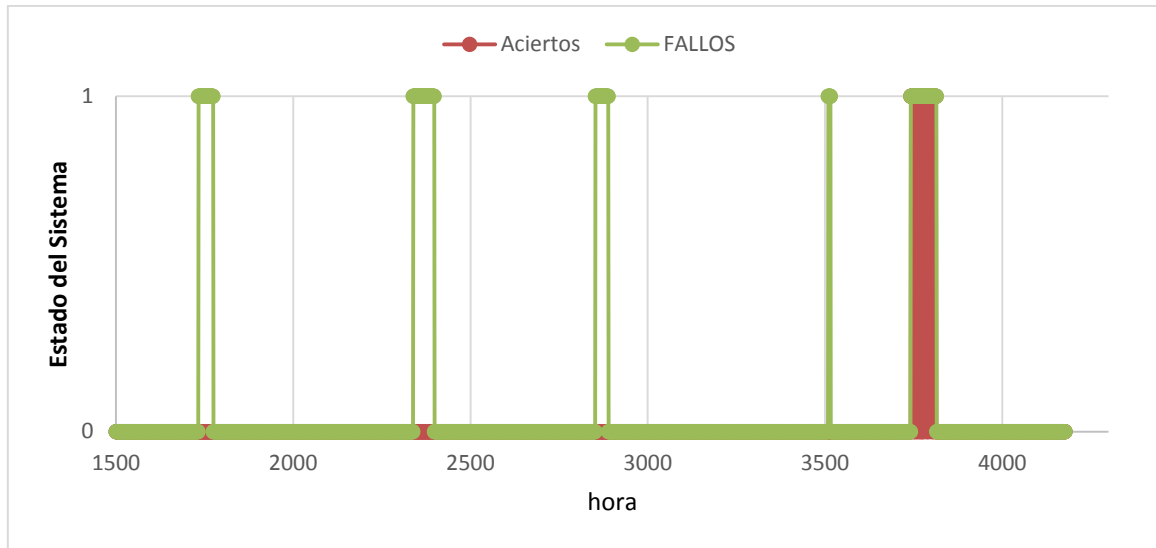


Figura 65. Localización de Aciertos Patrón 3

Por último en el Sector 5, el patrón ha detectado los dos primeros periodos de fallos, aunque no completamente, pero con una mayor efectividad que los demás. Estos valores se los pudo observar también en la **Figura 46** cuando se mostró tres grupos de valores que estaban muy alejados de la disposición general. A continuación, se muestra la ubicación de las detecciones durante toda la simulación:

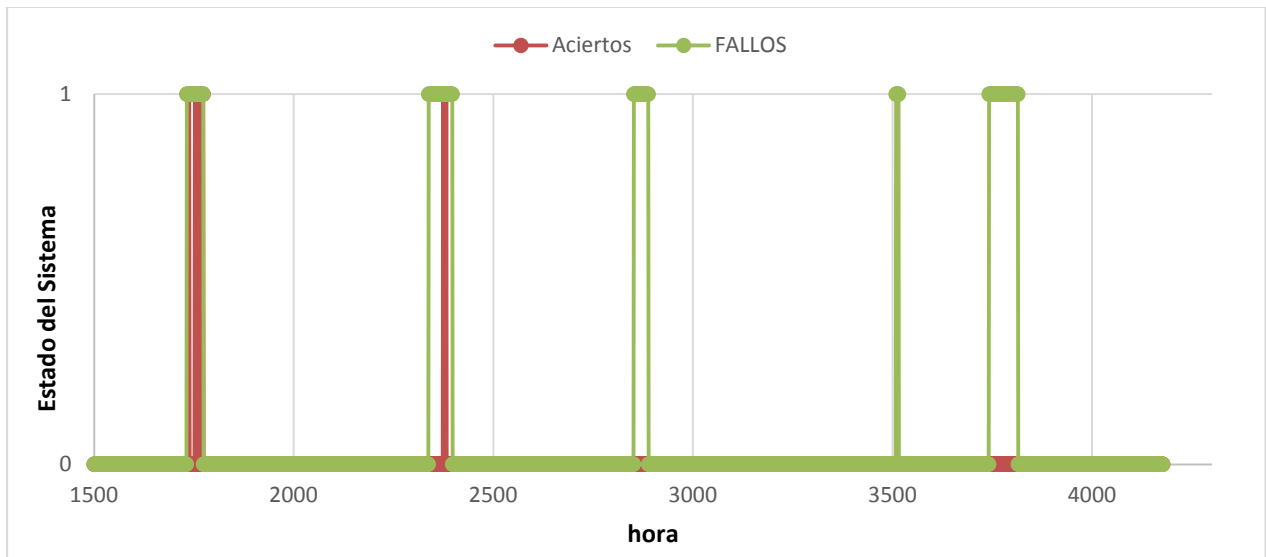


Figura 66. Localización de Aciertos Patrón 5

En principio es entendible estos porcentajes de detecciones, puesto que se está tratando de un valor de consumo que en muchas de las curvas de modulación tienen rangos amplios y entendiendo la naturaleza del consumo que varía de sector a sector, se obtuvo resultados medianamente buenos como era de esperarse. Por otro lado, también es importante resaltar que todos los periodos han

sido detectados, aunque no en la totalidad de los instantes, pero efectivamente el criterio reconoce valores demasiado sospechosos, que en muchos casos, hacen parte de varios periodos de fallos de la red.

5.2 ANOMALIAS EN LAS BOMBAS.

Como se miró en el capítulo 3, las estaciones de bombeo fueron descritas a partir de la elaboración de las curvas características, desarrollando así un criterio de semejanza en el que los datos del dataset 2 debían estar dentro de los rangos de funcionamiento para no ser clasificados como sospechosos. Entendido esto, a continuación, se presentará la cantidad de datos sospechosos que fueron encontrados producto de la evaluación del dataset 2, así como la efectividad de este criterio frente a la cantidad de alarmas encontradas. La **Tabla 19** recoge esta información a continuación:

DESCRIPCION	BOMBAS							
	PU1	PU2	PU4	PU6	PU7	PU8	PU10	PU11
SOSPECHOSOS	7	9	0	80	39	11	44	32
ACIERTOS	7	8	-	59	10	5	30	24
EFFECTIVIDAD	100%	88.8%	-	73.8%	25.64%	45.5%	68.2%	75%

Tabla 19. Resultados de la detección por bombeo.

Como podemos apreciar, los valores encontrados para este criterio de evaluación son mejores en comparación al criterio de consumos, con un promedio total del 70 % en cuanto a la efectividad de detección. En este mismo sentido, por ejemplo, para la bomba PU4 no se detectó ninguna anomalía, por lo cual no se pudo identificar ningún acierto, y esto es comprensible porque según la curva característica de la **Figura 50**, todos los valores eran muy similares a los que se encontró en el dataset 1.

También es importante mencionar que existen valores muy elevados de detección en la mayoría de las bombas, teniendo el valor más alto la bomba PU1 con el 100% de efectividad a la hora de detectar fallos en la red. De esta manera, para poder apreciar de una mejor manera que periodos de anomalías han sido detectadas, se procederá a representar la simulación completa de estos periodos junto con las detecciones.

Para la bomba PU1 como se observa en la **Figura 67**, el modelo ha detectado solo en el segundo periodo de fallos, que aunque son pocos, todos coinciden con las anomalías en la red. Por otro lado, considerando lo encontrado en el capítulo 4, cuando se analizó el dataset 2, se observaba en la **Figura 48** a manera de deducción, los 7 valores que no estaban dentro del rango de funcionamiento,

que aunque se ajustaban perfectamente a la curva característica, mostraban un comportamiento inhabitual porque funcionaban a caudales más bajos de lo normal.

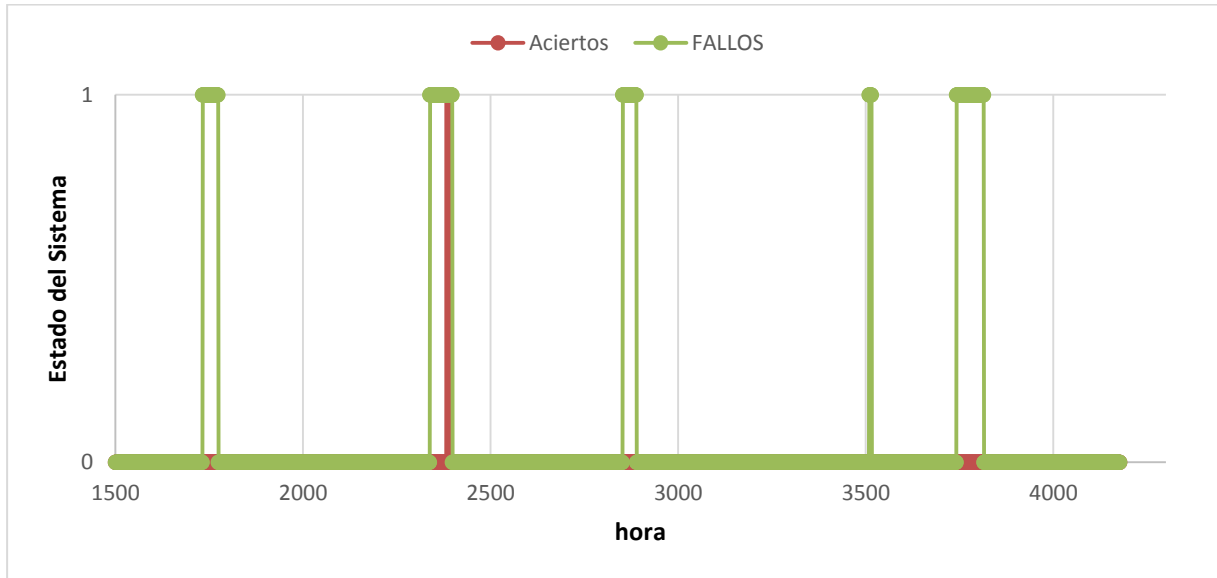


Figura 67. Localización de aciertos bomba 1

En cuanto a la bomba PU2, las detecciones son similares a la bomba anterior, puesto que se hicieron en el periodo 2 de fallos, y con una alta efectividad. Igualmente en la **Figura 49**, en principio se observó estos valores que estaban alejados del rango de funcionamiento normal, por lo cual era evidente que serían fallos en el sistema. La **Figura 68** muestra la disposición de las detecciones durante la simulación:



Figura 68. Localización de Aciertos bomba 2

Siguiendo con el análisis de resultados, la bomba PU4 no detecto ninguna anomalía y esto era evidente porque la mayoría de los valores mostrados en la **Figura 50** correspondían a simple vista a un comportamiento normal del sistema. Por otro lado, la bomba PU6 que pertenece al Sector 3, ha detectado muchos instantes de los últimos dos periodos de fallos como se muestra en la **Figura 69**, donde esta bomba ha localizado la mayoría de los fallos en el periodo 5 y algunos del periodo 4 teniendo así una alta efectividad de detección.

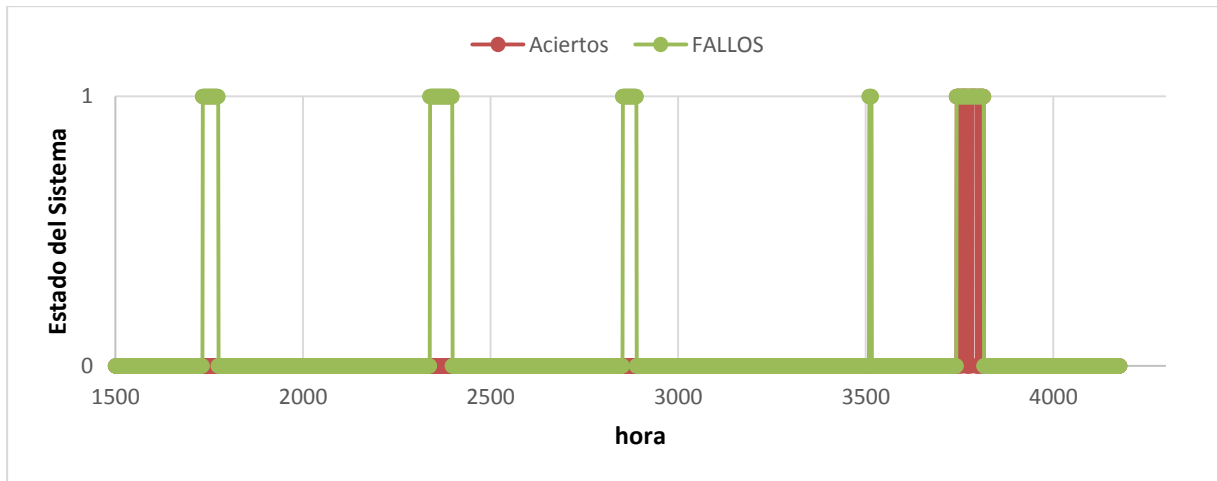


Figura 69. Localización de Aciertos bomba 6

En este mismo sector se encuentra la bomba PU7, la cual posee la más baja efectividad de todas, pero que aun así ha detectado 10 instantes en el último periodo de fallos, concluyendo evidentemente que la anomalía se basaba en la alteración de la velocidad de giro de la bomba observada en la **Figura 52**, donde se apreciaba hasta tres curvas características distintas para una misma bomba. La **Figura 70** muestra la localización de los aciertos durante la simulación.

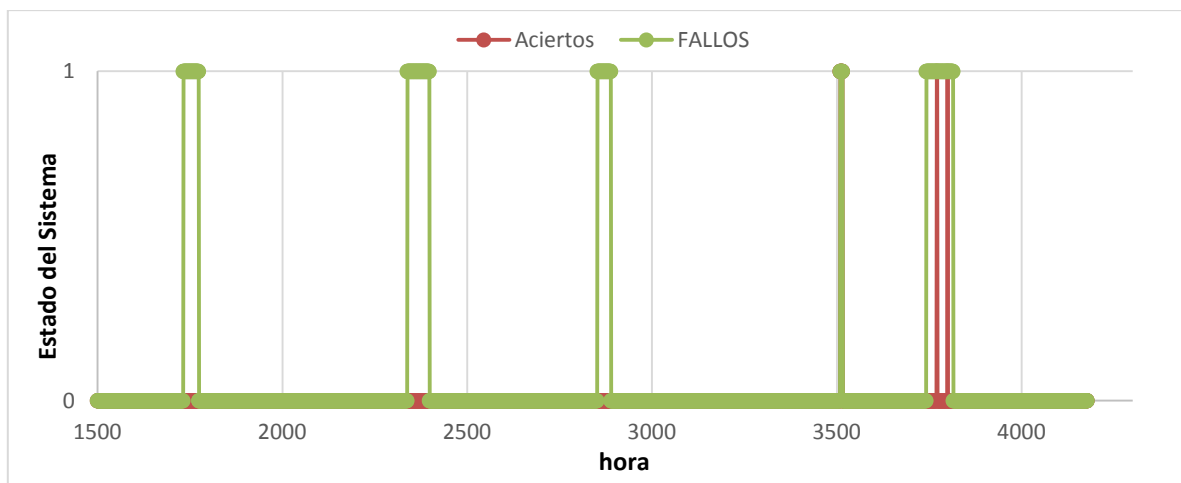


Figura 70. Localización de Aciertos bomba 6

Para la bomba PU8, como se observa en la **Figura 71**, la localización de las detenciones se llevaron a cabo en el periodo 2 y 3 de fallos, confirmando efectivamente lo que se había mostrado en la **Figura 54**, donde se mostraba claramente un par de valores que se salían del rango de funcionamiento, por lo cual la bomba funcionaba bien excepto por unos pocos instantes donde aumentaba el caudal inyectado de una forma no habitual.



Figura 71. Localización de Aciertos bomba 8

Entre tanto la bomba PU10 perteneciente al Sector 5, localiza la mayoría de los aciertos en el primer periodo de fallos, detectado casi la mayoría de los instantes y también algunos del periodo 2 y 3. Este resultado es coherente con lo visto en la **Figura 55**, donde evidentemente existían dos nubes de puntos por encima y debajo de la curva característica, en consecuencia, estos valores son en gran parte anomalías del primer periodo. A continuación, se presenta los aciertos del análisis de la bomba PU10.

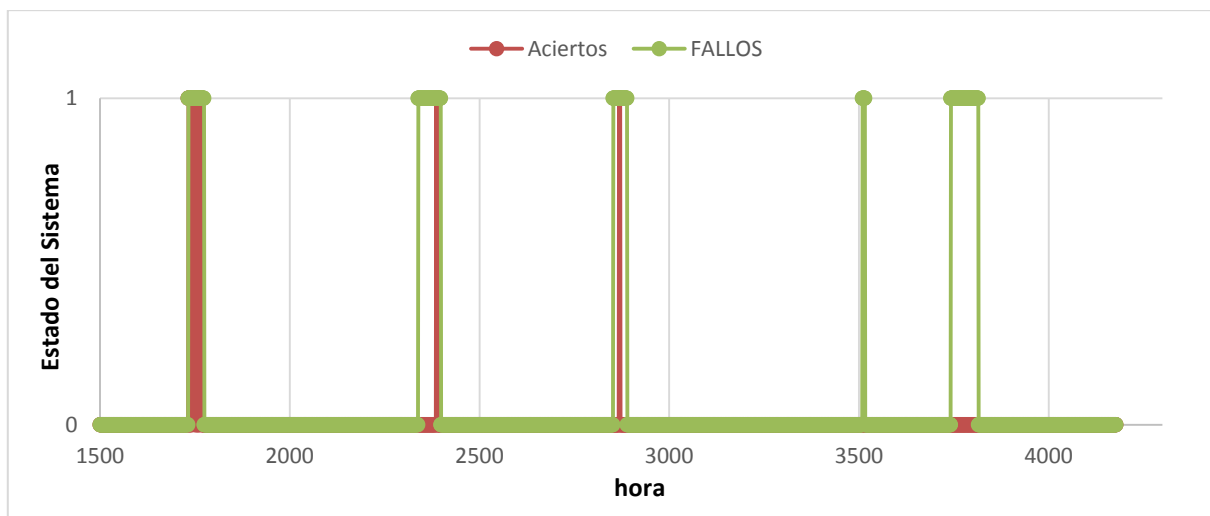


Figura 72. Localización Aciertos bomba 10

Por último, la bomba PU11 en principio no presenta detección alguna, puesto que como se miró en la **Figura 56** no existía alteración alguna a la curva característica de la bomba PU11 y su funcionamiento se limitaba a pocos instantes donde estaba encendida, sin embargo, esta bomba según el dataset 1, nunca estuvo en funcionamiento durante 1 año de funcionamiento, por lo cual se decidió por considerar este funcionamiento como sospechoso. Esta determinación ha sido muy acertada de acuerdo al porcentaje de efectividad que presenta, por lo cual se concluye que el encendido de la bomba PU11 siempre estuvo directamente relacionado con fallos en el sistema. A continuación, se muestra la localización de estas detecciones.

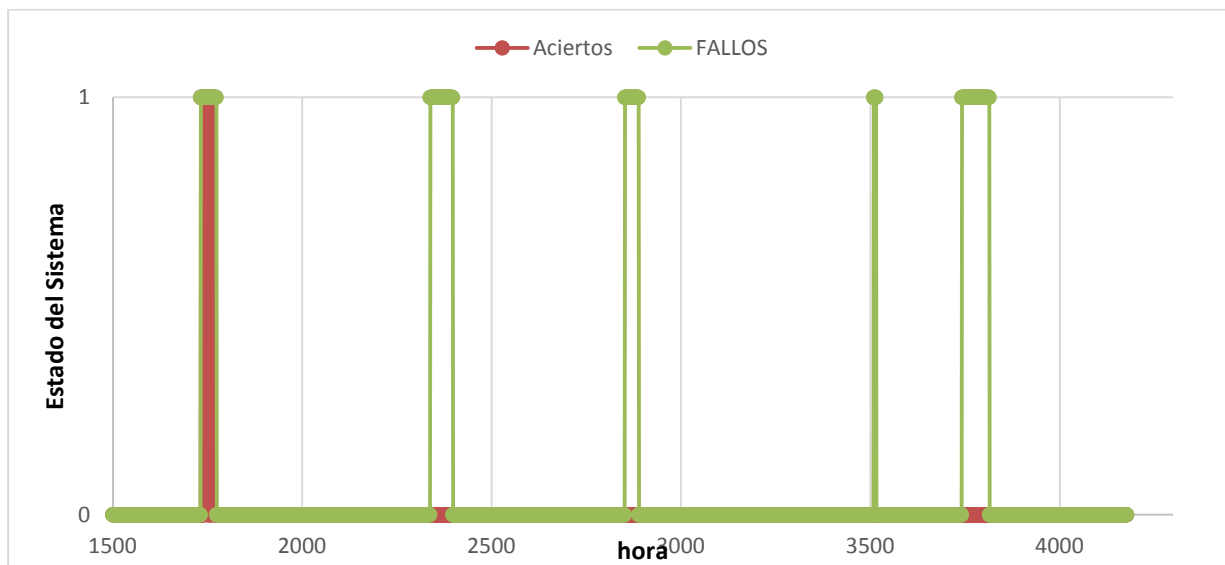


Figura 73. Localización Aciertos bomba 11

Como se observó durante toda la recolección de resultados por parte del criterio de bombas, esta forma de detección es muy efectiva, por lo cual hasta ahora ha sido la que más grandes resultado ha presentado, por encima de los consumos.

5.3 ANOMALIAS EN LOS TANQUES.

Para la recopilación de los resultados concernientes al funcionamiento de los niveles de los tanques, previamente se ha calculado los rangos de nivel correctos de acuerdo a los límites que cada uno presenta en condiciones normales. Con esto, se detectó diferentes anomalías en las cuales era evidente en algunos casos, comportamientos totalmente anormales; por lo cual se recogerá todos los datos sospechosos y se comparará con los periodos de fallo, de tal manera que se pueda identificar los aciertos que este criterio ha detectado durante toda la simulación. A continuación, se muestra los resultados de la comprobación de los datos en los tanques.

DESCRIPCION	TANQUES							
	T1	T2	T3	T4	T5	T6	T7	
SOSPECHOSOS	80	0	0	8	22	0	0	
ACIERTOS	24	-	-	5	3	-	-	
EFFECTIVIDAD	30%	-	-	62.5%	13.6%	-	-	

Tabla 20. Resultado de la detección en los tanques

La **Tabla 20** nos indica que en general la detección por anomalías en los niveles de los tanques ha sido medianamente eficiente, donde en el caso de del Tanque 1, se ha detectado muchos instantes de fallo a pesar de su eficiencia baja; por otro lado el tanque donde mejor se ajustó el criterio fue en el 4, donde se tiene la efectividad mayor del 62.5 %, aunque con muy pocos valores sospechosos y de aciertos. Así el Tanque 5 solo ha detectado un par de valores de fallo y los demás no han presentado datos sospechosos por lo cual en ninguno de estos últimos se presentan detecciones. Particularmente en el Tanque 1, las detecciones se han realizado en el periodo 2 de fallos, donde se ha detectado 24 fallos, los cuales se muestran en la **Figura 74**.

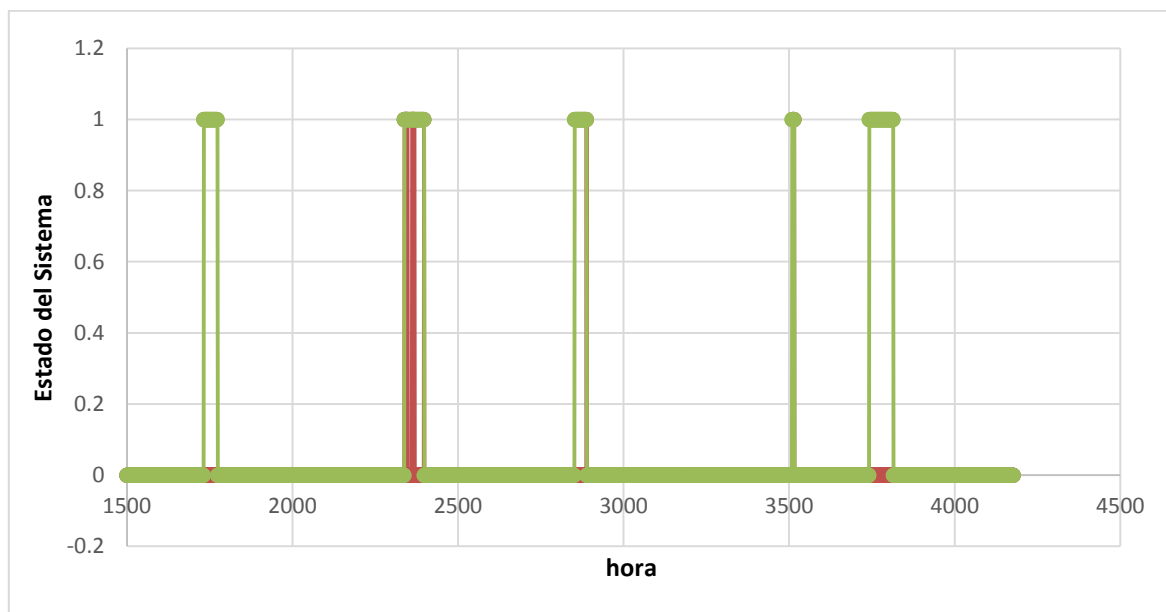


Figura 74. Localización Aciertos Tanque 1

El Tanque 4 por su parte, ha detectado 5 instantes del periodo 4 de fallos, mientras que el Tanque 5 tan solo 3 detecciones en el último periodo. Por otro lado, es importante mencionar que este criterio es el que menos resultados ha producido hasta ahora en cuanto a la detección de anomalías, sin embargo, existen muchos instantes que solo han sido detectados por este criterio, por lo cual, aunque no ha sido muy eficiente, la suma total de aciertos para la Matriz de resultados, depende significativamente de este criterio.

5.4 MATRIZ DE RESULTADOS

Como dijimos en el segundo capítulo, atendiendo a la forma de evaluación que BATADAL ha propuesto como criterio de evaluación final, en este punto del proyecto, se recogerá toda la información concerniente a los resultados del modelo y se analizarán los resultados desde un punto de vista más global, de forma que se puedan obtener números finales de detecciones y sus respectivas clasificaciones según la Matriz de resultados.

Antes de conocer los valores de la Matriz de resultados, es importante mencionar la cantidad de detecciones que el modelo ha logrado identificar a partir de cualquiera de los criterios usados, es decir, el modelo al menos una vez ha logrado identificar un instante y lo ha clasificado como sospechoso. En este sentido en la **Tabla 21** se muestra el periodo de fallo junto con la duración en horas y así mismo la cantidad de instantes que el modelo ha detectado.

PERIODO DE FALLO	DURACIÓN (h)	DETECCIONES (h)	% DETECCION
1	42	37	88.1
2	60	31	53.3
3	37	12	35.1
4	7	7	100
5	73	72	98.6

Tabla 21. Números generales de detección.

Analizando los resultados obtenidos, se encuentra que existen varios periodos que han sido detectados casi en su totalidad, por ejemplo, el periodo 5 ha sido detectado casi completamente; también el periodo 4, donde su detección es el 100%. Por otro lado, el periodo que menos instantes ha sido detectado es el 3, donde se tiene un porcentaje de solo el 35%, seguido por el periodo 2 con el 53.3%.

MATRIZ DE RESULTADOS		Estado Actual	
		FALLO (positivo)	SEGURO (Negativo)
Estado Predicho	FALLO (positivo)	159	199
	SEGURO (negativo)	60	3759
TOTAL HORAS		4177	

Tabla 22. Matriz de resultados del Modelo de detección



En este contexto, para determinar finalmente la matriz de resultados, se ha clasificado cada uno de las detecciones de acuerdo al tipo de característica descrito en el capítulo 2, de esta manera, se encontró una totalidad de 358 instantes donde al menos una vez el modelo lo clasifico como sospechoso; de esos, 159 instantes fueron fallos reales, es decir, que coincidieron con alguno de los periodos de fallos. Por otro lado, se encontró también 199 falsos positivos, es decir, el modelo considero que eran anomalías, pero en la realidad solo se trataba de simples sospechas. Por último, se encontraron 60 instantes donde el modelo no reconoció los fallos, que produce al final un total de 219 horas de anomalías para las cuales solo se detectó el 72% de ellas. A continuación, en la **Tabla 22** se muestran los valores para la matriz de resultados.



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA

máster en ingeniería
hidráulica y medio ambiente
mihma

6. MEJORAS DEL MODELO.

Después de la elaboración del modelo, de la comprobación del mismo y obtención de resultados, se ha intentado mejorar la efectividad del procedimiento, de modo que se pueda obtener el 100% de las detecciones. Para ello se ha puesto la mirada en el análisis de las variables de forma individual, de manera que se pueda entender el comportamiento durante todas las horas de registro y obtener información necesaria que lleve a la detección de fallos. Estas mejoras se desarrollaron a partir del análisis de las siguientes variables:

- Presiones a la entrada y salida de las estaciones de bombeo y válvula.
- Flujo individual de las bombas.
- Caudal de la válvula.

Con el estudio detallado de estas variables se espera conseguir los resultados más óptimos posibles, de tal forma que la red C-Town pueda estar protegida totalmente de las anomalías que puedan ser provocadas o del resultado de algún tipo de accidente. A continuación, se desarrollará el análisis de cada uno de estos puntos y se presentará al final los resultados en a través de la Matriz de resultados.

6.1 ANALISIS DE LAS PRESIONES.

Como se sabe, los registros de la red cuentan con mediciones directas de la presión en la entrada y salida de las estaciones de bombeo. Por eso, usando estas mediciones se analizará las variaciones de presión durante toda la simulación, usando evidentemente comportamientos normales del dataset 1 y comparando como se hizo durante el modelo con los resultados del dataset 2. Para el análisis de las presiones, se desarrollará sector a sector, detallando el comportamiento de sus respectivas mediciones y subrayando por lo tanto las mejoras encontradas. El criterio que se usará para evaluar el dataset 2 será el siguiente:

- Todos aquellos valores que estén por fuera del rango normal de funcionamiento encontrado en el dataset 1 serán marcados como sospechosos.

Con la implementación de este criterio, se espera igual que en los registros de los tanques, detectar las nubes de puntos que no hagan parte de la tendencia normal del Sistema y así entender de una mejor manera el funcionamiento de las estaciones de bombeo respectivas.

6.1.1 PRESIONES ESTACION DE BOMBEO 1.

Comenzando con este análisis, el Sector 1 como se observó en la Figura 16, posee la medición de los puntos J280 y J269 que corresponden a la entrada y salida respectivamente. En cuanto al primero, la **Figura 75** muestra un comportamiento bastante claro de la entrada, donde los valores son próximos a 3 mca y no existen variaciones considerables. También es importante mencionar la existencia de franjas de funcionamiento, las cuales se pueden apreciar a simple vista, de modo que cada franja tendrá su valor máximo y mínimo para su posterior comparación.

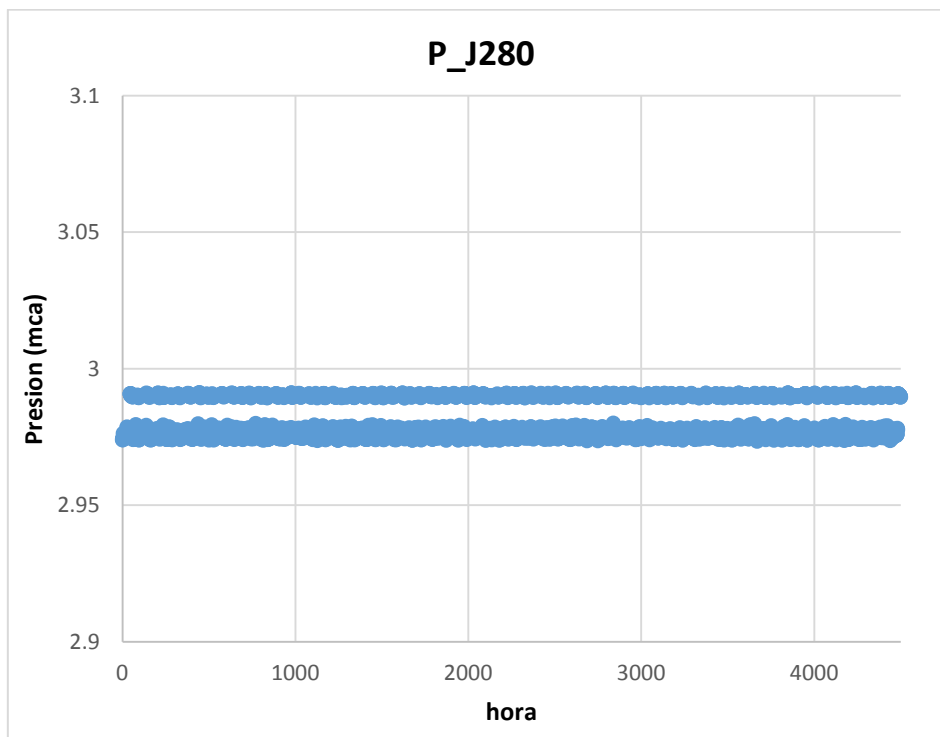


Figura 75. Presiones a la entrada de la Estación de bombeo 1 dataset 1.

En cuanto a la salida de la estación de bombeo, el dataset 1 nos muestra dos franjas de presiones, que se puede entender a partir del funcionamiento de las bombas, puesto que los valores mayores son el resultado del bombeo por parte de una sola bomba, en este caso de la bomba PU1, mientras que los valores inferiores corresponden al funcionamiento en paralelo con la bomba PU2. Este planteamiento se lo puede observar en la **Figura 76**.

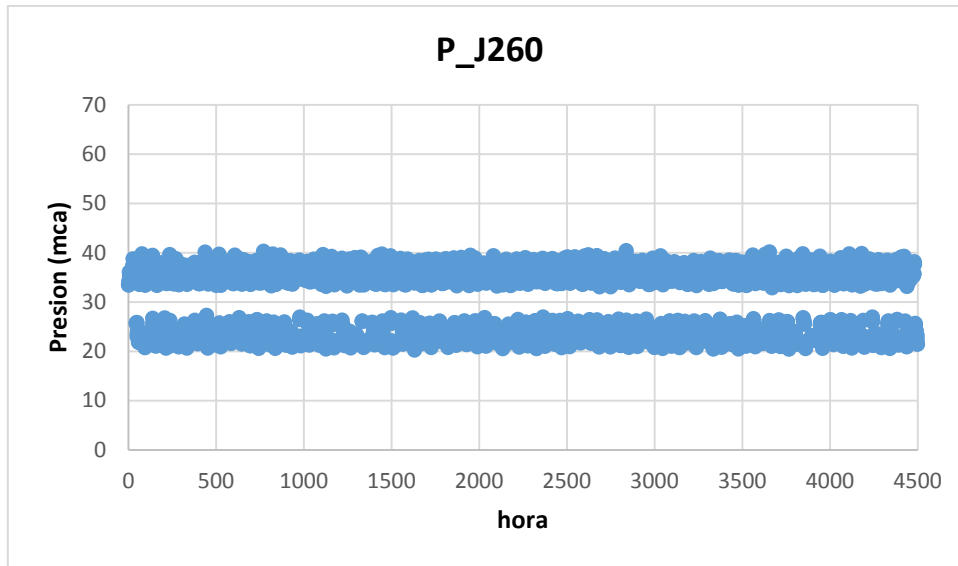


Figura 76. Presiones a la salida de la Estación de bombeo 1 dataset 1.

Descrito el comportamiento normal de las presiones en la entrada y salida, es necesario observar la disposición de los valores para el dataset 2. En la **Figura 77** se observan valores muy semejantes a los observados en el dataset 1 a diferencia de un punto que está por encima del rango superior y que al parecer concuerda con el segundo periodo de anomalías. De esta manera se aplica el criterio propuesto de modo que los valores fuera de la tendencia correcta serán posibles anomalías.

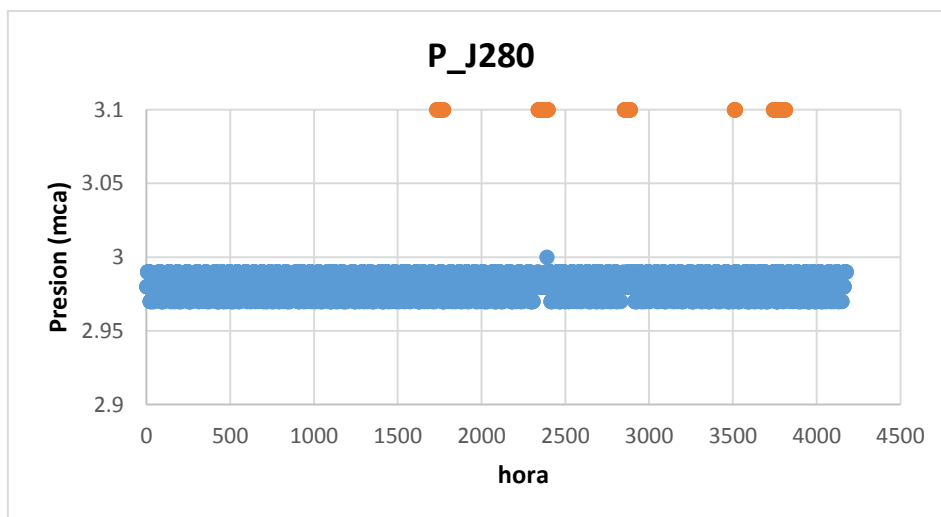


Figura 77. Presión a la entrada de la Estación de bombeo 1 dataset 2.

Para los valores de la salida de la estación de bombeo 1 en el dataset 2, se ve claramente en la **Figura 78** valores sospechosos entre la hora 2000 y 2500, que precisamente coinciden con el segundo periodo de anomalías, por lo cual existirán varios valores que harán parte de las sospechas para este análisis.

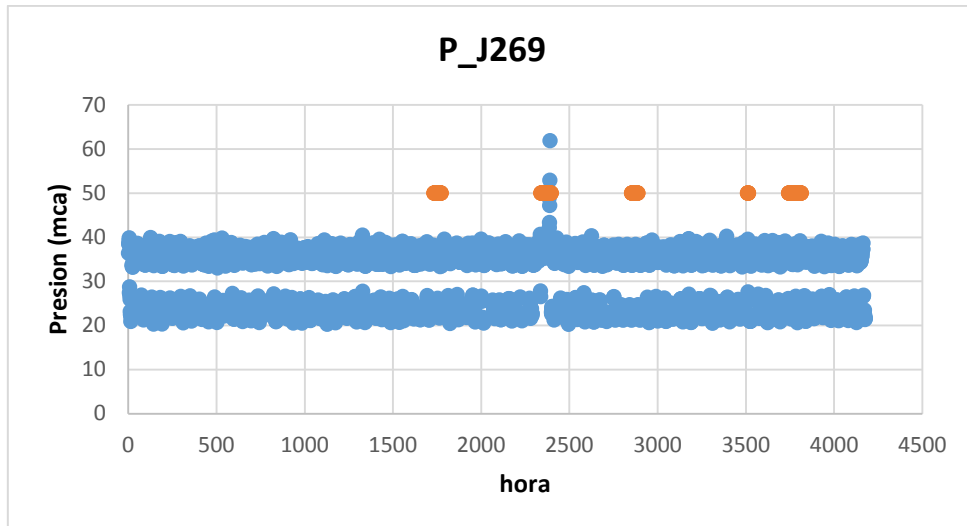


Figura 78. Presión a la salida de la Estación de bombeo 1 dataset 2.

6.1.2 PRESIONES ESTACION DE BOMBEO 2.

Continuando con el análisis, el Sector 2 cuenta con la medición del punto J300 y J256 como se puede observar en la **Figura 79**. Hay que recordar que en este sector solo funcionaba la bomba PU4 y la bomba PU5 se mantenía apagada durante todo el tiempo de registro de los datos, en consecuencia, solo se aplicara el criterio a la bomba que está en funcionamiento.

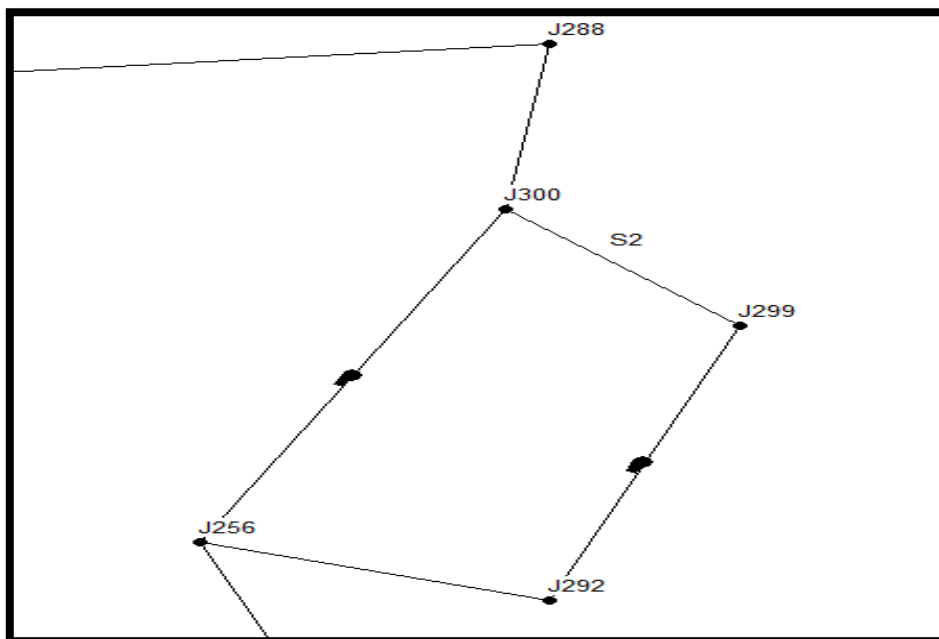


Figura 79. Estación de bombeo Sector 2

Teniendo claro esto, a partir del dataset 1 se encontró para la estación de bombeo 2 una tendencia marcada, donde los valores oscilan entre 20 y 34 mca como se observa en la **Figura 80**. Por lo cual, a partir de estos valores máximos y mínimos se aplicará la evaluación para el dataset 2.

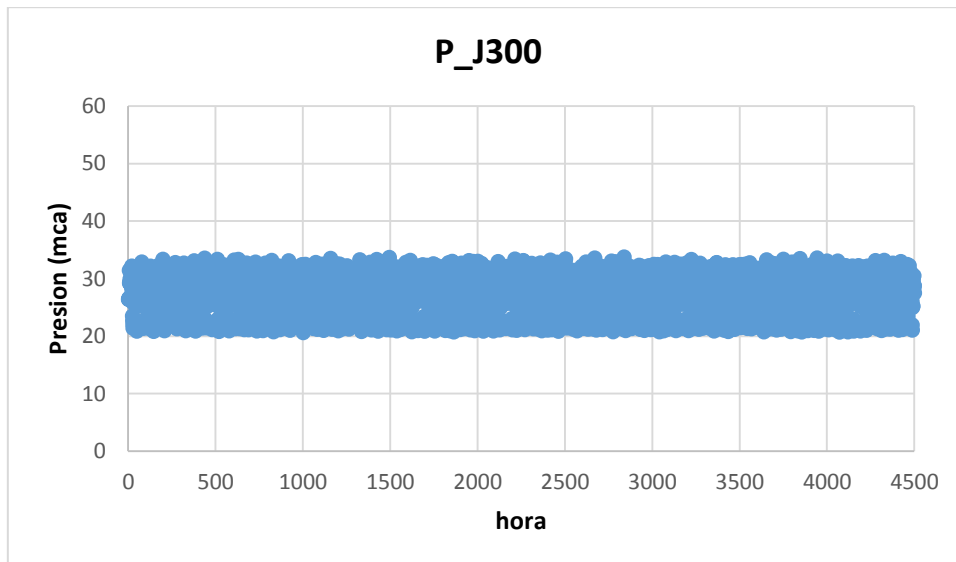


Figura 80. Presión a la entrada de la estación de bombeo 2 dataset 1

Para la salida de la estación de bombeo 2, se encontró que al igual que en la salida del Sector 1, existen dos rangos de funcionamiento con algún grado de variabilidad en ellos, pero que de igual forma se calculara sus valores máximos y mínimos para compararlos con los registros del dataset 2. A continuación se presentan los datos de la salida del Sector 2 durante todo el periodo de medición en el dataset 1.

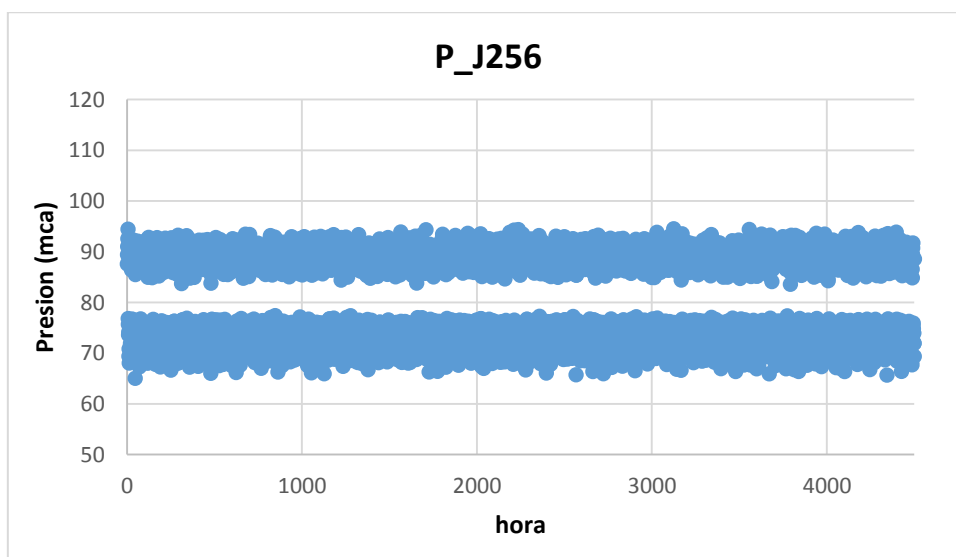


Figura 81. Presión a la salida de la Estación de bombeo 2 dataset 1

Teniendo claro el comportamiento normal del sistema, ahora se analizará las presiones del dataset 2 y como se observa en la **Figura 82**, en principio no existe gran variabilidad de los datos, sin embargo algunos valores que se alejan de la tendencia coinciden con los periodos de fallo del sistema, por ejemplo, entre la hora 2500 y 3000.

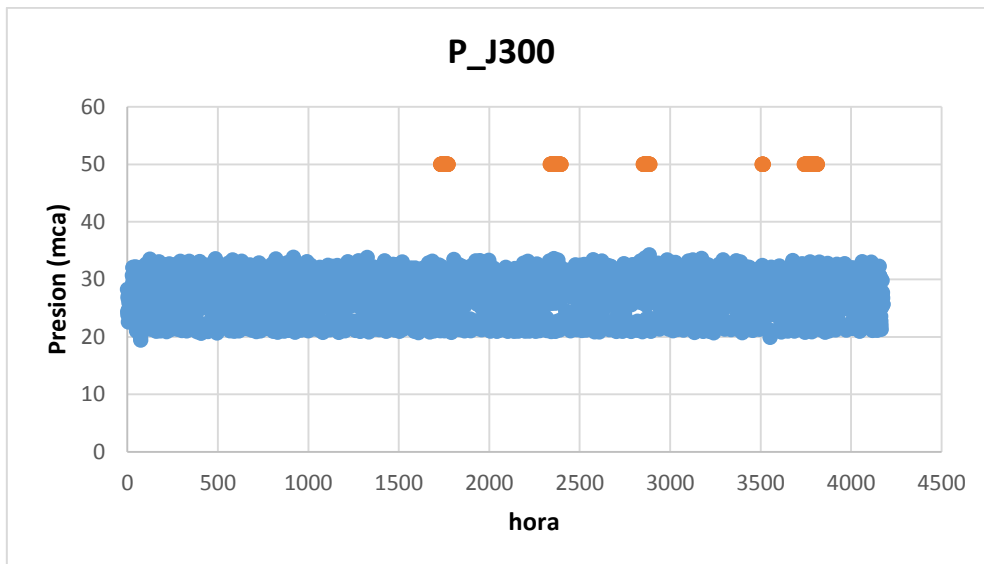


Figura 82. Presión a la entrada de la Estación de bombeo 2 dataset 2

De la misma manera los datos de salida de la estación de bombeo 2 para el dataset 2, se muestran en la **Figura 83**, donde se ve igual que en el dataset 1 dos rangos de valores, los cuales se validaran a partir del criterio establecido, de forma que todos los registros que estén fuera de los rangos normales de funcionamiento serán considerados sospechosos.

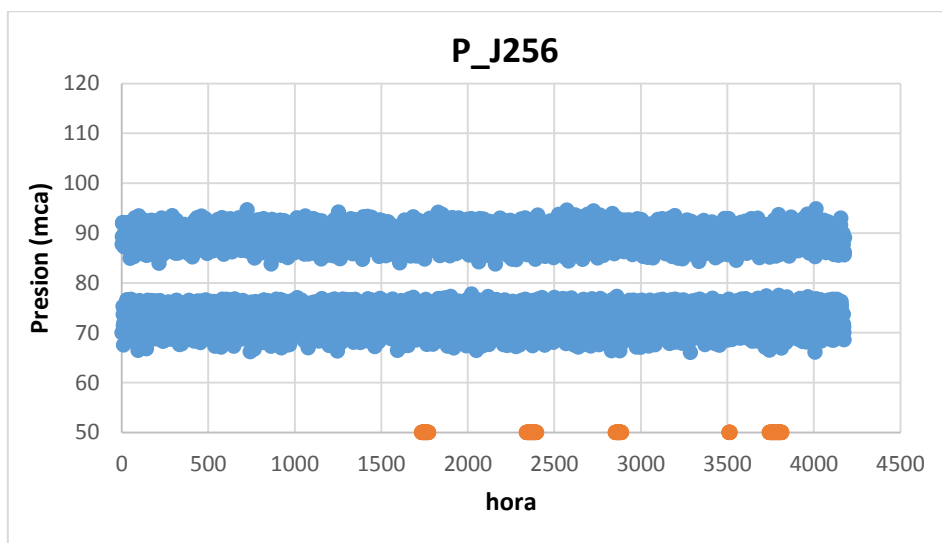


Figura 83. Presión a la salida de la Estación de bombeo 2 dataset 2

6.1.3 PRESIONES ESTACION DE BOMBEO 3.

A continuación, el Sector 3 contiene también los datos de presión de la entrada y salida de la estación de bombeo, es decir, las medidas del punto J289 y J415 respectivamente, como se observa en la **Figura 84**.

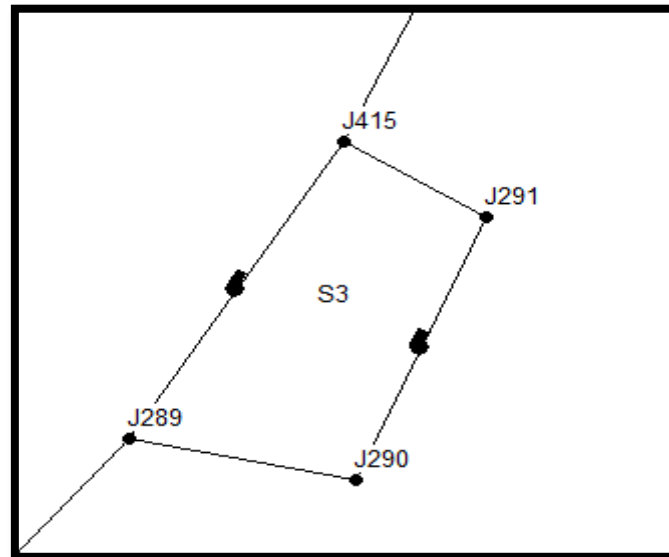


Figura 84. Estación de bombeo Sector 3.

Analizando el dataset 1, se encuentra que para la entrada de la estación de bombeo 3, los valores están bien definidos y mantienen una tendencia entre los 20 y 34 mca como se observa en la **Figura 85**. Este comportamiento es muy parecido al encontrado en el Sector 2 para el punto J300.

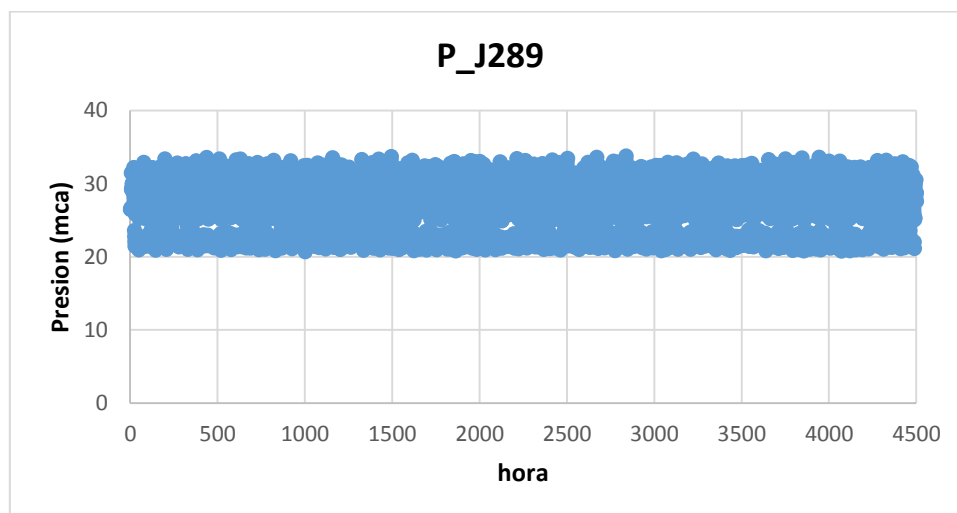


Figura 85. Presión a la entrada de la Estación de bombeo 3 dataset 1

Continuando con el Sector 3, la salida de la estación de bombeo presenta una variabilidad mayor en comparación a los anteriores datos encontrados, por lo cual el punto J415 tendrá rangos de funcionamiento mayores. Por otro lado, también es notorio la existencia de dos rangos de funcionamiento, por lo cual cada uno de ellos tendrá un valor máximo y mínimo que se contrastará más adelante con los del dataset 2.

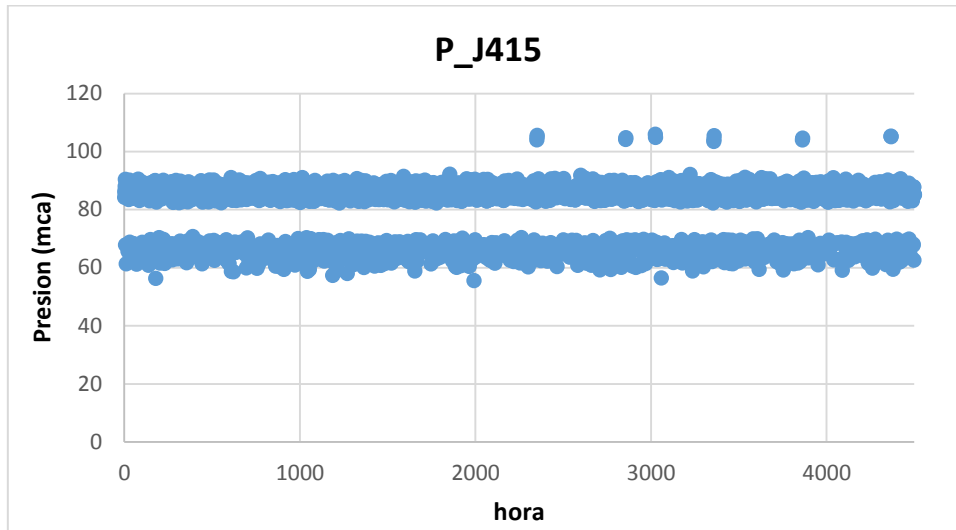


Figura 86. Presión a la salida de la Estación de bombeo 3 dataset 1.

Entendido el funcionamiento correcto del Sector 3, se ha de analizar la entrada y salida para el dataset 2. En cuanto a la presión de entrada, se tiene una tendencia parecida a la encontrada en el dataset 1 como se ve en la **Figura 87**, sin embargo, se aplicará el criterio a partir de los límites correctos encontrados.

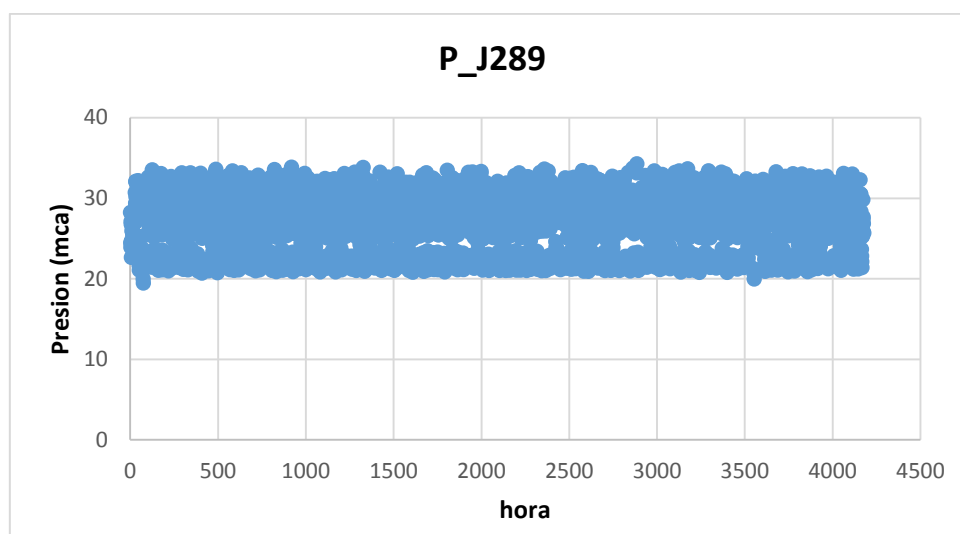


Figura 87. Presión a la entrada de la Estación de bombeo 3 dataset 2.

Igualmente, en la salida de esta estación se encontró una disposición de los registros bastante sospechosa como se observa en la **Figura 88**, donde gran cantidad de datos están muy distantes de la tendencia habitual. Por ejemplo, pasada la hora 3500, existe una agrupación de valores que coinciden con el periodo 4 de fallos, de forma que estos pueden ser probablemente anomalías del sistema. Por último al igual que todos los sectores, se evaluarán conforme al criterio establecido y se etiquetarán todos aquellos registros que se salgan del comportamiento habitual como sospechosos.

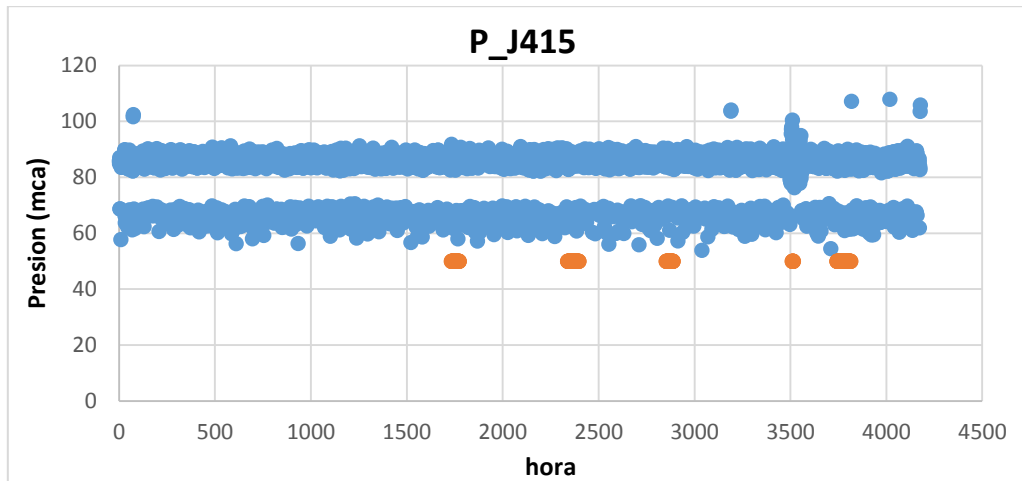


Figura 88. Presión a la salida de la Estación de bombeo 3 dataset 2

6.1.4 PRESIONES ESTACION DE BOMBEO 4.

En cuanto al Sector 4, los puntos que corresponden a la entrada y salida de la estación de bombeo son el J302 y el J306 como se muestra en la **Figura 89**. Por lo cual, se mostrará a continuación el comportamiento normal del sistema para su posterior comparación.

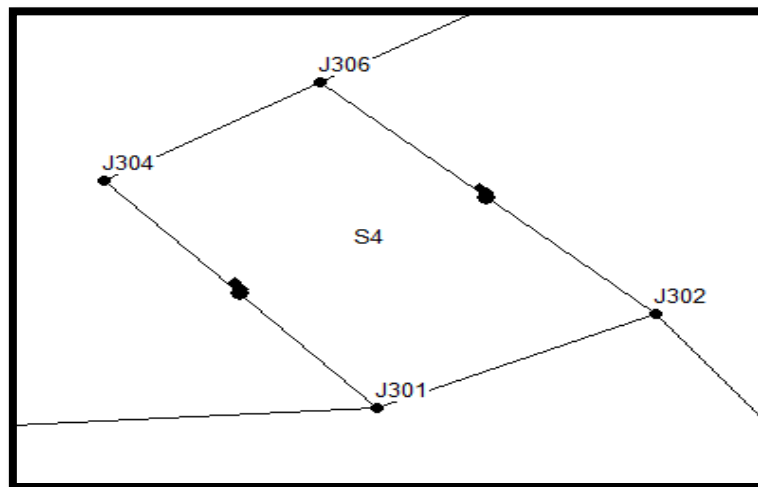


Figura 89. Estación de bombeo Sector 4

Para la entrada de la estación de bombeo, en el punto J302 se obtuvo dos franjas de valores que tienen cierto grado de variabilidad, por lo cual se determinan los límites para esta variable y de modo que se pueda establecer luego la comparación respectiva.

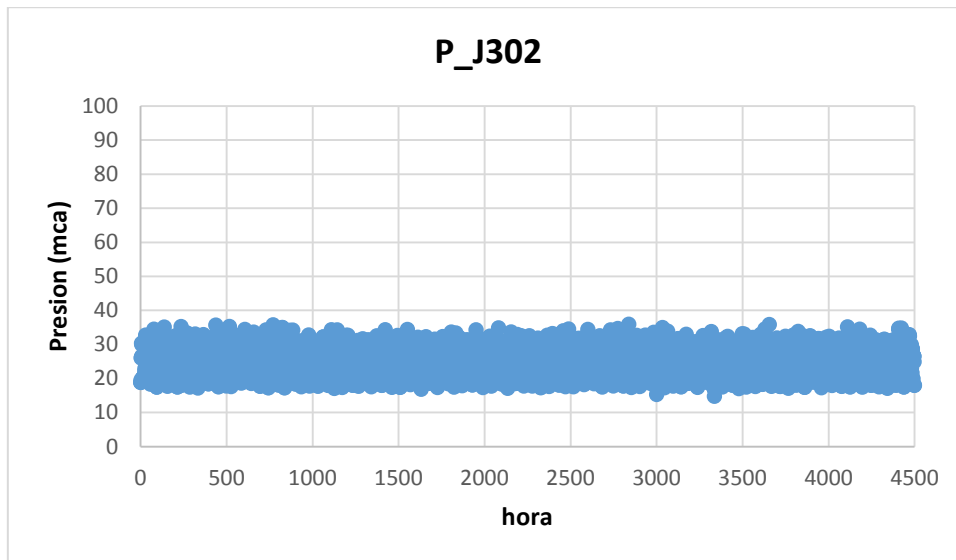


Figura 90. Presión a la entrada de la Estación de bombeo 4 dataset 1

Del mismo modo, para la salida de la estación de bombeo, notamos en la **Figura 91** la existencia de dos franjas de funcionamiento y en cierto modo con un pequeño grado de variabilidad. Por lo cual se definen los límites de funcionamiento normal de presión para la salida a partir del cálculo de los niveles máximos y mínimos de los rangos encontrados.

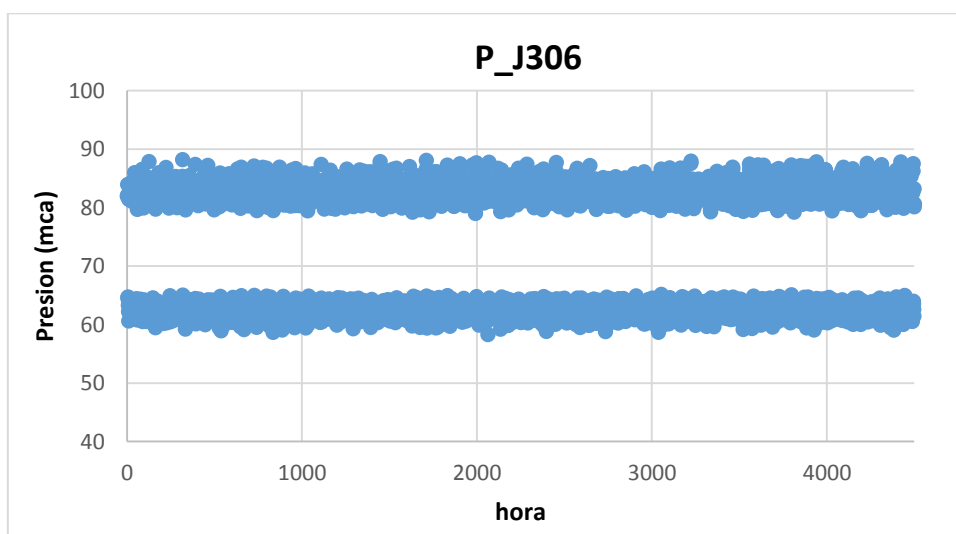


Figura 91. Presión a la salida de la Estación de bombeo 4 dataset 1.

Siguiendo con el procedimiento, ahora es necesario observar los registros del dataset 2, los cuales contienen las anomalías en el sistema. En la **Figura 92** se observan muchos registros sospechosos, que en la mayoría de casos coinciden con los periodos de fallo en la red. En este caso coinciden los datos sospechosos con los periodos 2 y 3 de anomalías, por lo cual se aplicará el criterio establecido para seleccionar estos valores como sospechosos.

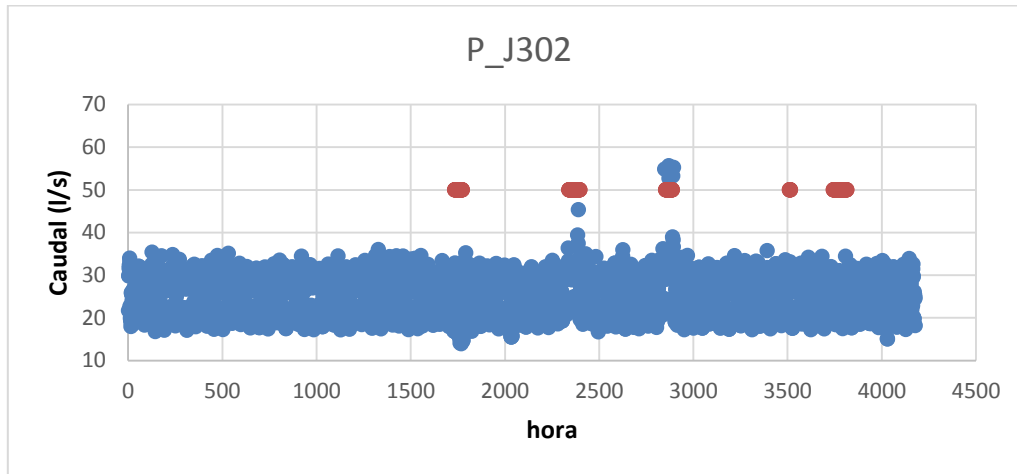


Figura 92. Presión a la entrada de la Estación de bombeo 4 dataset 2

Para terminar con este sector, la salida de la estación de bombeo tiene un comportamiento igualmente sospechoso como se observa en la **Figura 93**, el cual también presenta registros anormales al comportamiento general. Precisamente estos valores coinciden con los rangos de fallos, sobre todo con el segundo y tercer periodo. De este modo, todos aquellos que no hagan parte del rango establecido a partir del dataset 1, serán marcados como sospechosos.

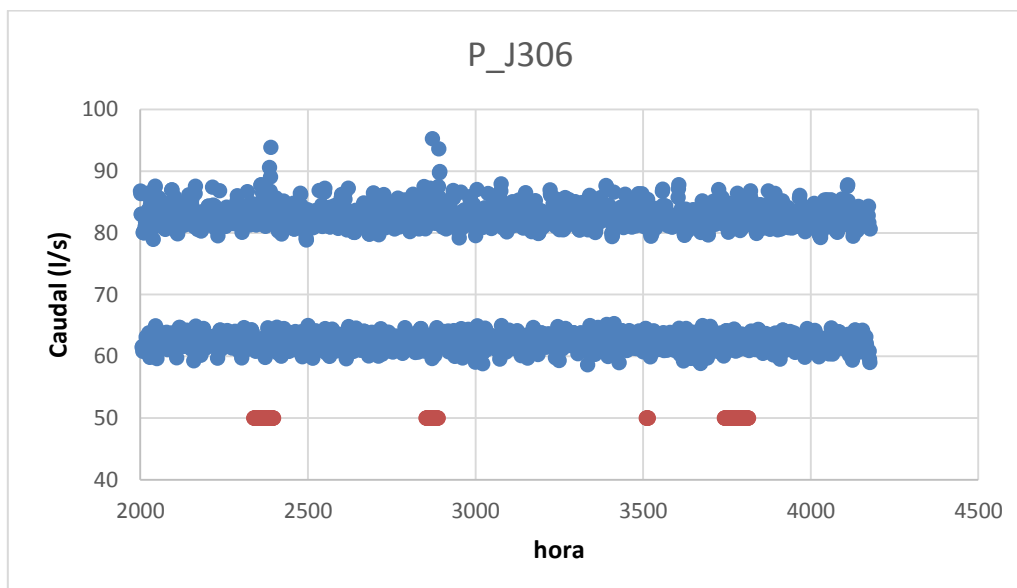


Figura 93. Presión a la salida de la Estación de bombeo 4 dataset 2

6.1.5 PRESIONES ESTACION DE BOMBEO 5.

Por último, el Sector 5 que se muestra en la Figura 94, cuenta también con los registros de los puntos J307 y J317. De esta forma, igual que en los demás sectores se usará el dataset 1 como base de comprobación para los registros.

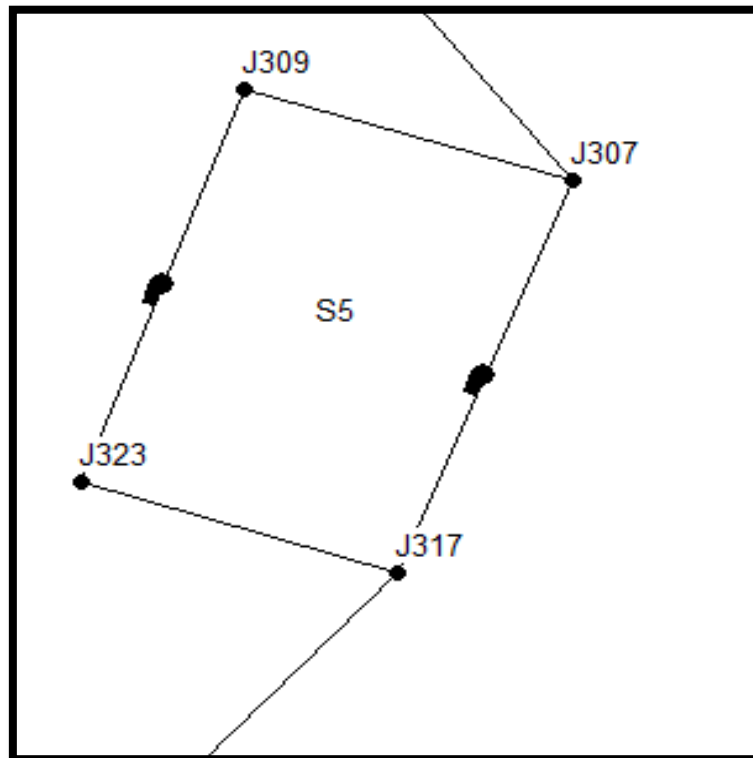


Figura 94. Estación de bombeo 5

Empezando el análisis en este sector, se encontró que para la entrada de la estación de bombeo 5, la disposición de los registros tiene dos rangos de valores como se observa en la **Figura 95**, que están muy unidos entre sí pero que se alcanzan a distinguir a simple vista. Por otro lado, también es importante mencionar que existe variabilidad en los registros, sobre todo en el límite superior, donde se nota claramente valores un tanto dispersos de la tendencia normal del caudal. De este modo se fija los valores máximos y mínimos de la entrada como puntos de partida para la posterior evaluación del dataset 2.

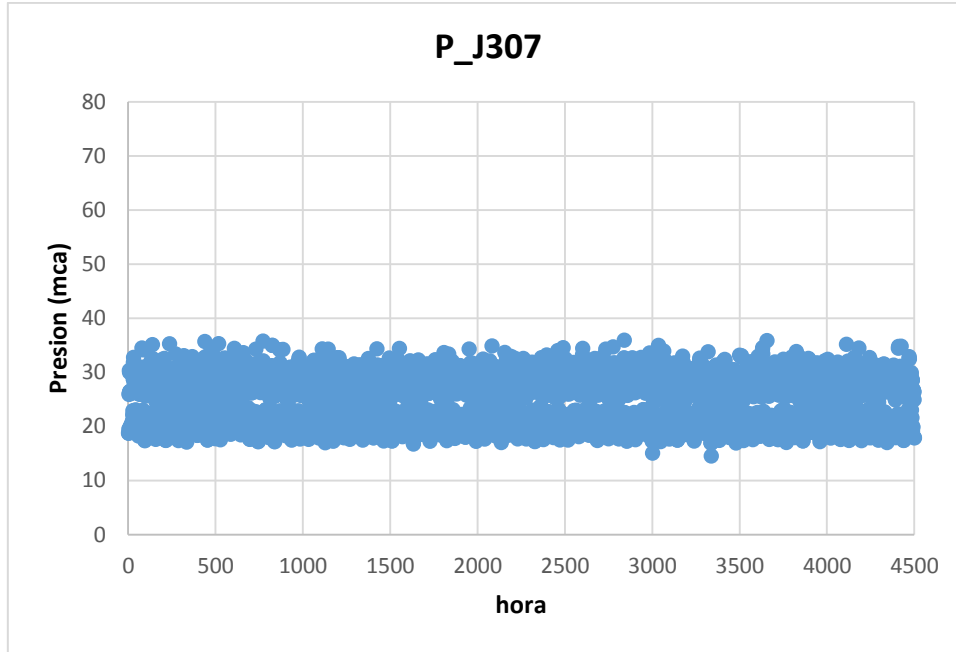


Figura 95. Presión a la entrada de la Estación de bombeo 5 dataset 1

A continuación, en la **Figura 96** se indica la tendencia de los valores durante todo el tiempo de captación datos, arrojando cierta variabilidad en algunas horas pero en general, con una tendencia marcada por dos regiones de datos.

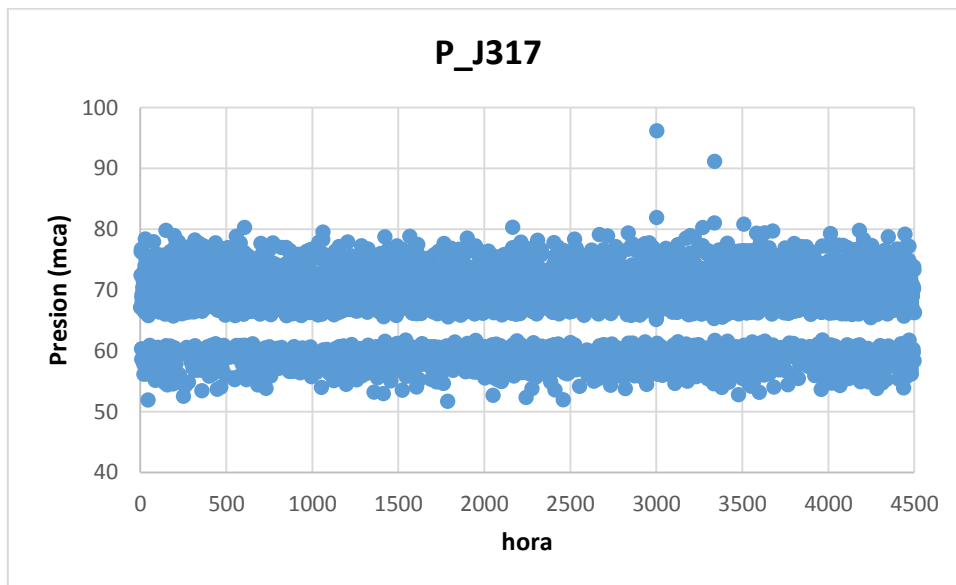


Figura 96. Presión a la salida de la Estación de bombeo 5 dataset 1.

Así, con el comportamiento normal establecido, se analizará el dataset 2 para este sector. En cuanto a la entrada se nota claramente varios valores que no coinciden a ninguno de los rangos, por lo cual aplicando el criterio de evaluación se apartaran todos los valores que estén por fuera de los límites como posibles anomalías del sistema. La siguiente Figura muestra el comportamiento de la entrada de esta estación para el dataset 2.

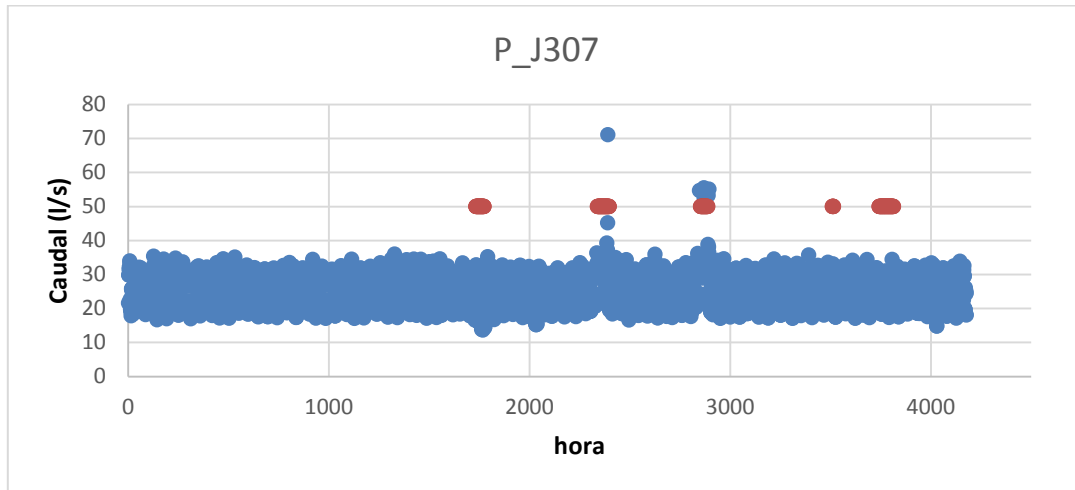


Figura 97. Presión a la entrada de la estación de bombeo 5 dataset 2

Así mismo, en la salida notamos muchos más registros sospechosos, por ejemplo, entre la hora 1500 y 2000, existe una nube de puntos muy variables en comparación a los demás que coincide con el primer periodo de anomalías, por lo cual muy posiblemente se traten de anomalías.

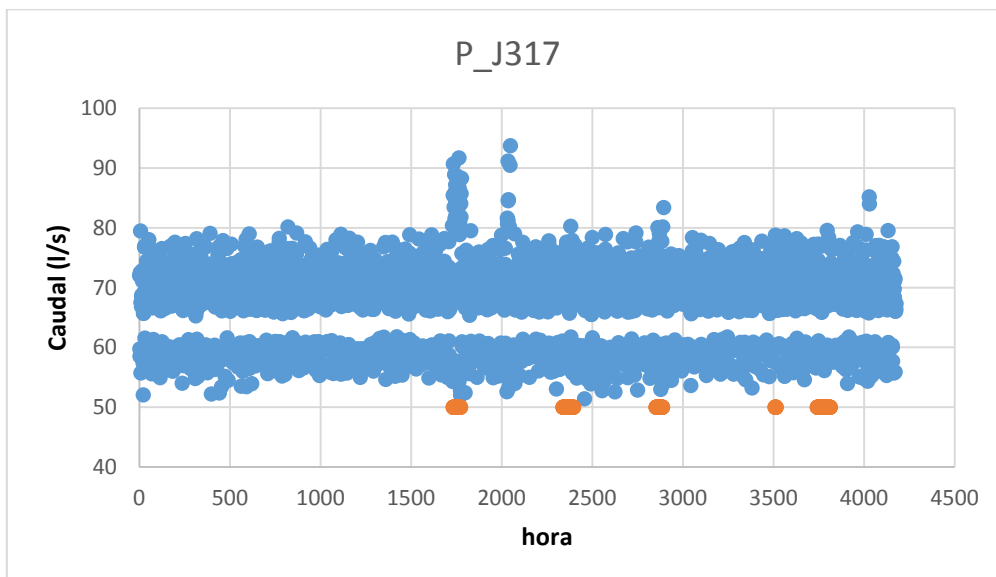


Figura 98. Presión a la salida de la Estación de bombeo 5 dataset 2

6.1.6 PRESIONES VALVULA DE REGULACION.

Terminado el análisis de las presiones en las estaciones de bombeo, se consideró la posibilidad de analizar el comportamiento de la válvula situada justo antes del Tanque 2 como se muestra en la **Figura 99**, de la cual se cuenta con registros de presiones en los puntos J14 y J422. En principio se había descartado el análisis de este elemento, pero a partir del análisis de las otras presiones, también se consideró la de esta válvula.

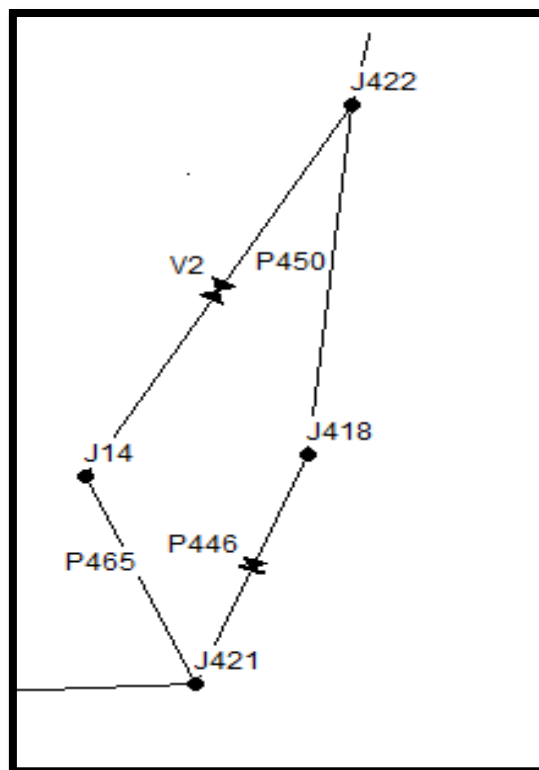


Figura 99. Válvula de Regulación

Partiendo del análisis de la entrada a la válvula, se ha encontrado una tendencia bien definida, con valores que van desde los 27 hasta los 46 mca, en dos rangos distintos de funcionamiento como se observa en la **Figura 100**. Estos límites se usarán para evaluar más adelante los registros del dataset 2.

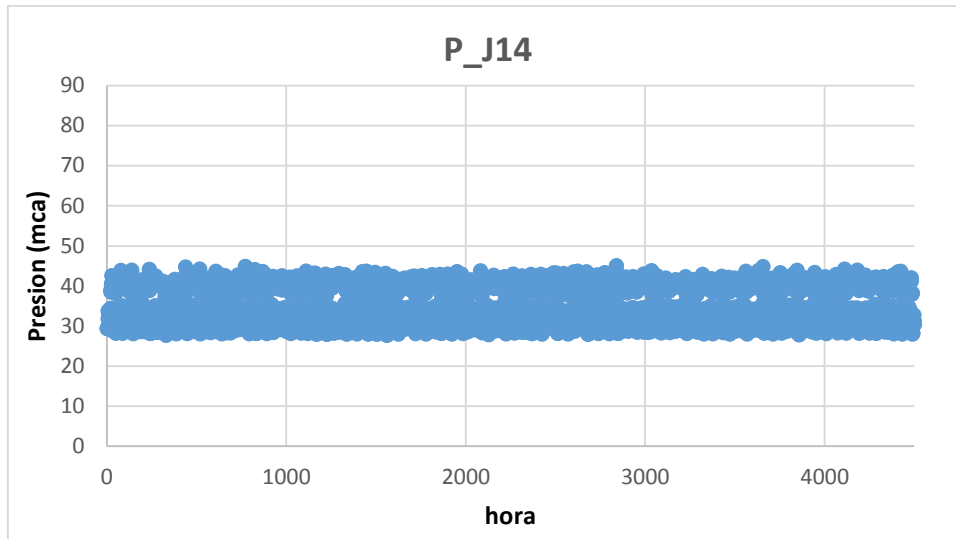


Figura 100. Presión a la entrada de la Válvula dataset 1

De esta misma manera, se analizó el comportamiento de la salida de la válvula para el dataset 1, encontrándose dos franjas de valores muy próximas entre sí, de las cuales se ha de calcular los valores máximos y mínimos para su respectiva comparación con el dataset 2. A continuación se muestra esta disposición para las presiones de salida.

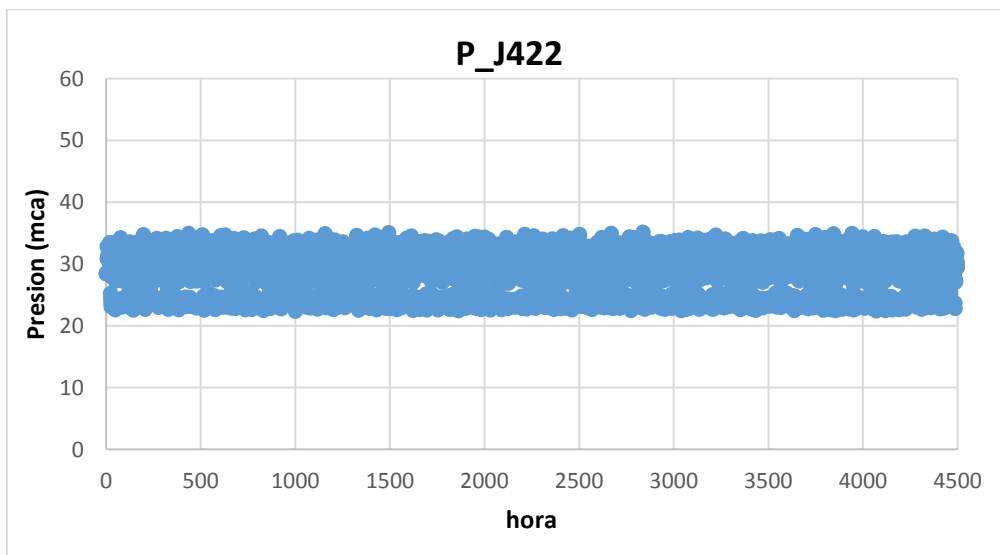


Figura 101. Presión a la salida de la Válvula dataset 1

Con esto en mente, se procede a analizar la válvula a partir del dataset 2, donde se encuentran diferentes instantes con registros bastante sospechosos, que en este caso coinciden con el periodo 2 y 3 de fallos como se aprecia en la **Figura 102**, por lo que se espera que muchos de estos valores efectivamente sean anomalías reales de la red.

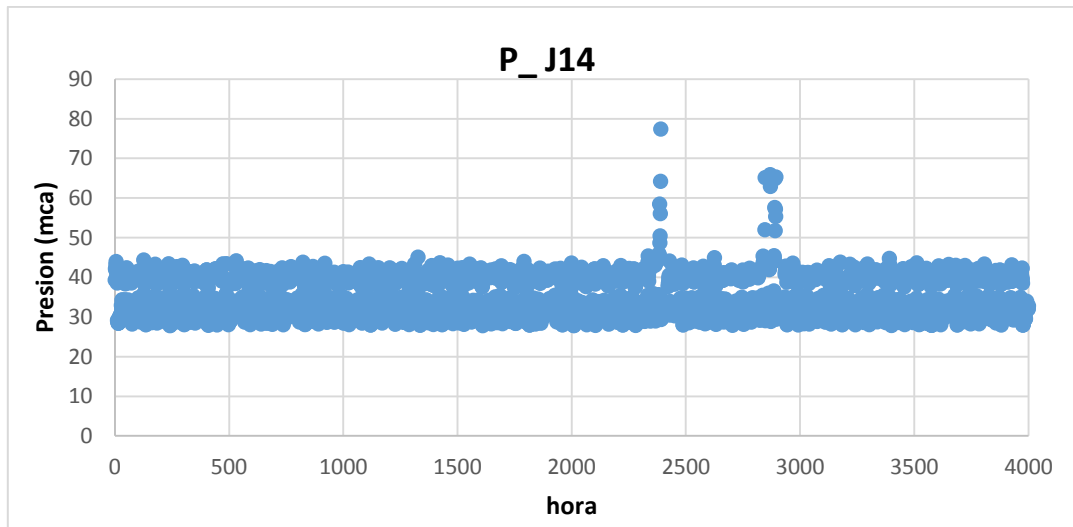


Figura 102. Presión a la entrada de la Válvula dataset 2.

Para el punto J422 correspondiente a la salida, se encontró una tendencia bastante parecida a la del dataset 1 como se observa en la **Figura 103**, en consecuencia, serán pocos los valores que se puedan determinar como sospechosos, dada la disposición de los registros durante todo el tiempo de simulación.

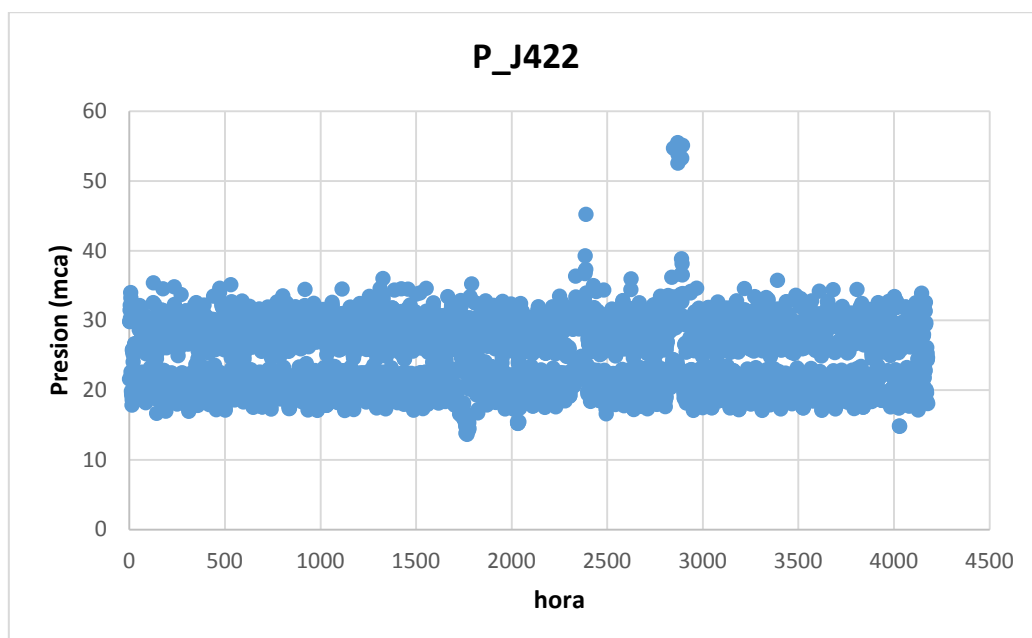


Figura 103. Presión a la salida de la Válvula dataset 2

Entendiendo que, en el caso de las presiones, solo se habían utilizado para calcular la altura de bombeo en cada instante de tiempo, este planteamiento permitirá añadir más valores sospechosos a la lista de posibles anomalías de cada estación de bombeo, de manera que aquellas horas que no detecto el análisis a las curvas características de las bombas, puedan realizarse a través de este criterio.

6.2 ANALISIS DE CAUDAL BOMBEADO

Para la mejora del modelo, se planteó también la posibilidad de analizar el flujo de caudal bombeado en cada estación, que en el caso de las curvas características y los patrones de consumo, se usó en conjunto con otras variables como los niveles de los tanques y las presiones; por eso se pensó inicialmente en concentrar la atención solo en los caudales que cada bomba inyecta a la red, de forma que durante todo el tiempo de registro se puedan encontrar datos no coherentes de los que se pueda intuir una posible anomalía en la red.

En este sentido, se usará un criterio que al igual que las presiones en el ítem anterior, determine rangos de funcionamiento óptimos a partir del uso del dataset 1 para luego contrastar los resultados con el dataset 2 y de esta manera seleccionar registros sospechosos. El criterio para la evaluación del caudal bombeado será el siguiente:

- Todo registro de caudal en la bomba respectiva que no se ajuste al rango encontrado a partir del análisis del dataset 1 será marcado como sospechoso.

De esta forma, se establece el criterio que como se mencionaba anteriormente, parte en principio del análisis de los registros del dataset 1 y terminara con el uso del mismo como base para la evaluación del criterio, de modo que todos estos valores etiquetados al final de la mejora se sumaran a los datos sospechosos del análisis de los patrones de consumo, debido a que está directamente relacionado con la veracidad del caudal que consume cada Sector. De esta manera, los valores sospechosos de caudal, provocarían dudas sobre el consumo real, de modo que se determinara al final, si realmente se está consumiendo todo el caudal inyectado o si simplemente se trata de un fallo en el funcionamiento de las bombas.

6.2.1 ANALISIS DEL CAUDAL BOMBEADO EN EL SECTOR 1

Es necesario recordar que el Sector 1 posee tres bombas, de las cuales solo funcionan dos: PU1 y PU2. En cuanto a la primera, como se observa en la **Figura 104**, se ve claramente dos rangos de funcionamiento que concuerdan con los representados en la curva característica de la **Figura 17**. De esta forma, se establecen los niveles máximos y mínimos para su posterior comparación.

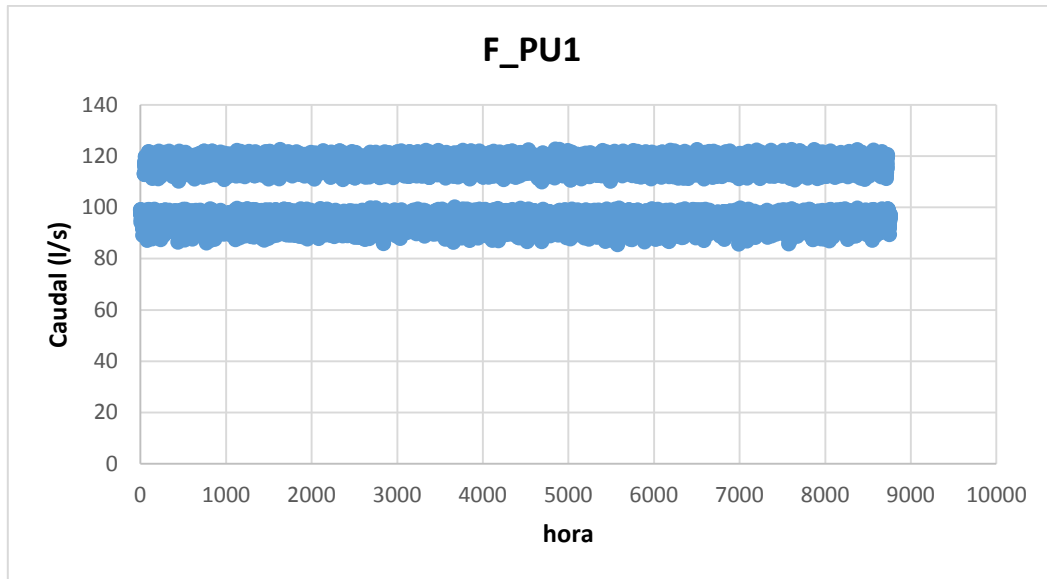


Figura 104. Caudal de la bomba 1 dataset 1

En cuanto a la bomba PU2, los datos mostrados en la **Figura 105** reflejan solo un rango de valores con cierto grado de variabilidad en cuanto al límite inferior. Hay que recordar también que esta bomba no siempre está encendida, de modo que estas variaciones pueden ser debido a estos arranques y paradas. En este sentido se establecen los valores límites superior e inferior de esta franja de valores como base de la evaluación del criterio.

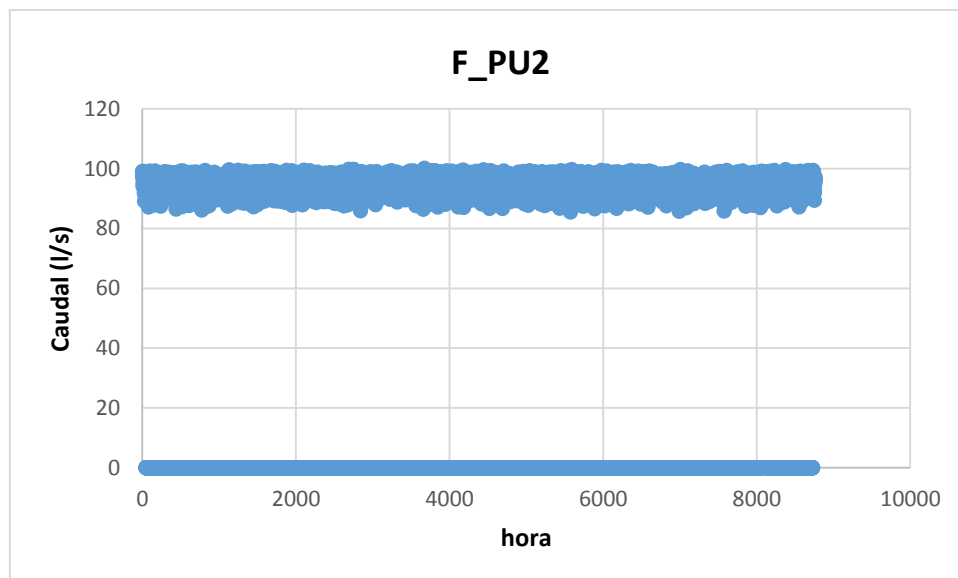


Figura 105. Caudal de la bomba 2 dataset 1

Luego de determinar los rangos de caudales correctos para estas bombas, se ha de analizar el dataset 2 en busca de posibles anomalías. En cuanto a la bomba PU1, se encontró muchos datos

sospechosos entre la hora 2000 y 2500 como se observa en la **Figura 106** y que concuerdan con el segundo periodo de fallos. Por lo cual estos valores harán parte de la lista de sospechosos a partir de la evaluación del criterio establecido.

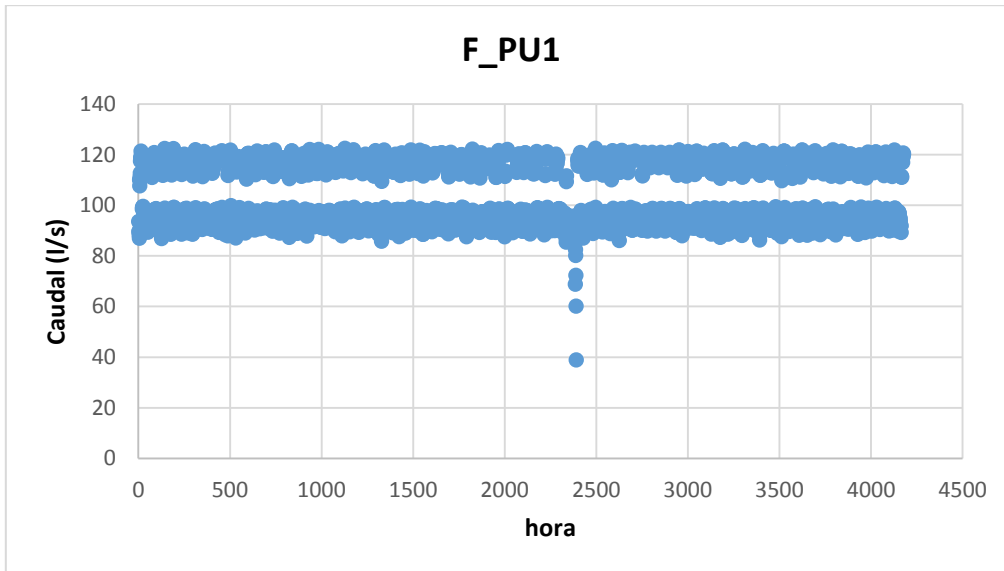


Figura 106. Caudal de la bomba 1 dataset 2

Por último, la bomba PU2 también presenta valores bastante sospechosos, y la disposición de los valores se parece a las de la bomba PU1, de modo que es evidente las anomalías en esta bomba. Para seleccionar estos valores, se aplicará igualmente el criterio de evaluación y se añadirán a la lista de sospechas del patrón 1.

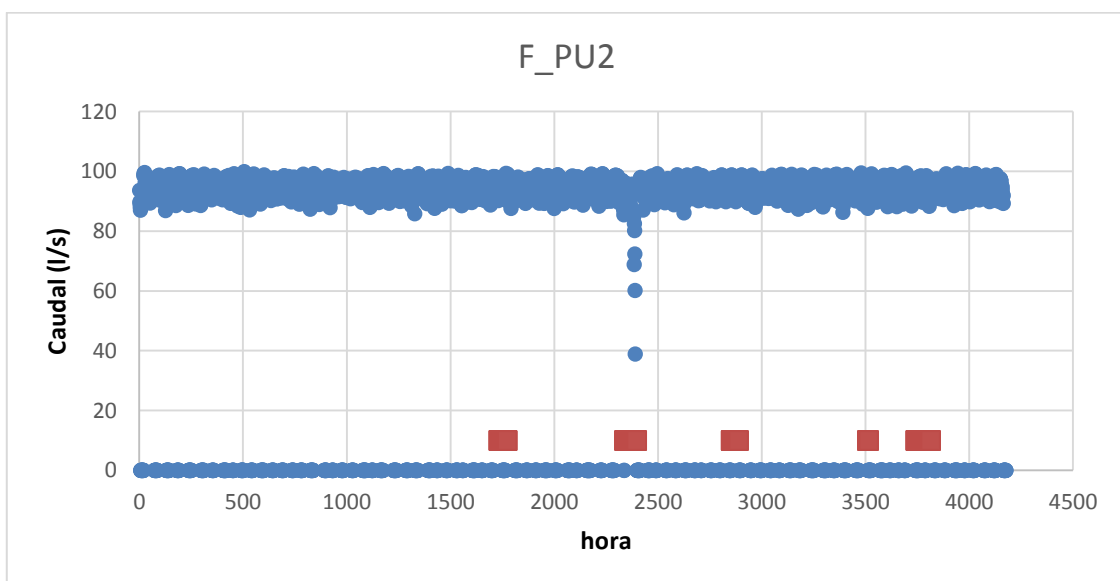


Figura 107. Caudal de la bomba 2 dataset 2.

6.2.2 ANALISIS DEL CAUDAL BOMBEADO EN EL SECTOR 2

Este sector posee dos bombas: PU4 y PU5; de las cuales solo funciona la primera y no en todos los instantes de tiempo, debido a que hay momentos donde la bomba está parada como se ve en la Figura 108, donde la franja inferior corresponde a los instantes donde está apagada. Por otro lado, también se ve un rango de valores cuando está encendida que va de los 30 a los 39 l/s aproximadamente.

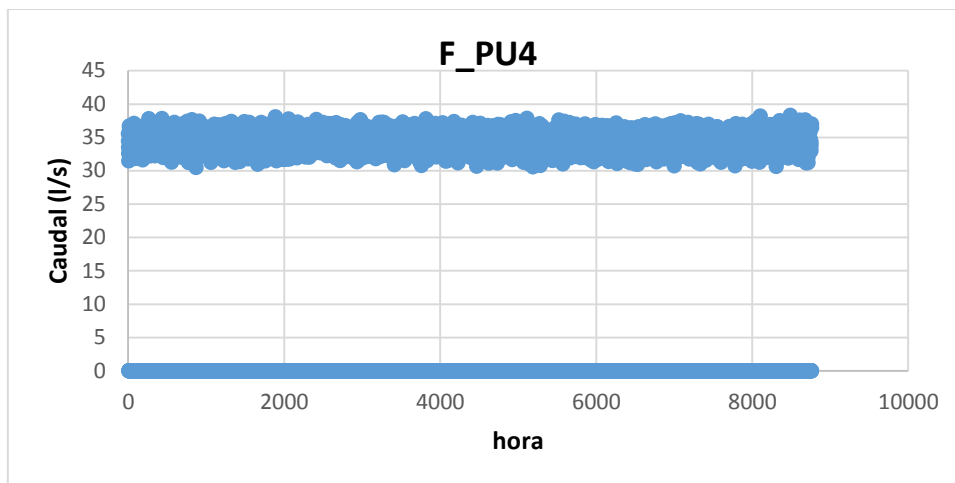


Figura 108. Caudal de la bomba 4 dataset 1

En cuanto al dataset 2, se encontró que para la bomba PU4, los valores parecían estar en orden como se ven en la Figura 109, encontrándose entre los rangos encontrados en el dataset 1 por lo cual a simple vista serán muy pocos o ningún valor que se pueda etiquetar como sospechoso.

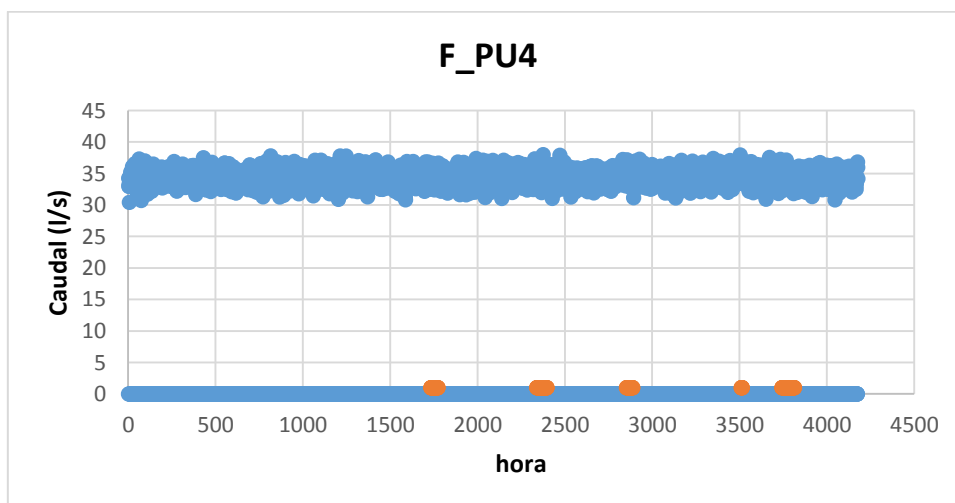


Figura 109. Caudal de la bomba 4 dataset 2

6.2.3 ANALISIS DEL CAUDAL BOMBEADO EN EL SECTOR 3

Para el Sector 2, al que pertenecen las bombas PU6 y PU7, se tiene que, para la primera, el caudal inyectado es cero, por lo cual solo se hará el análisis del caudal para la bomba PU7. De este modo, como se observa en la **Figura 110**, existen valores que se alejan de la tendencia normal del sistema, por eso los límites de este mismo serán mucho más amplios.

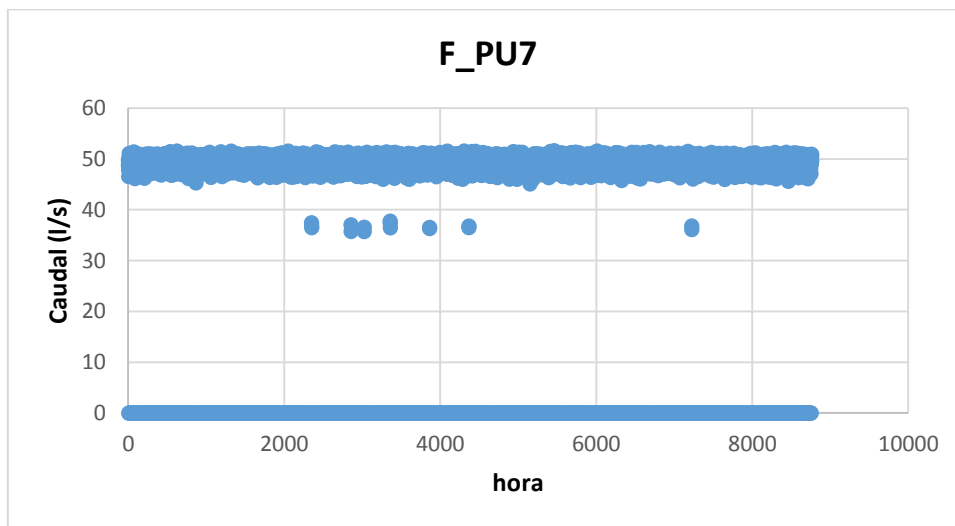


Figura 110. Caudal de la bomba 7 dataset 1

En este mismo sentido, se encontró que para el dataset 2, la bomba PU7 tiene instantes muy sospechosos como se ve en la Figura 111, donde los instantes de anomalías coinciden muy bien con estos. De modo que, aplicando el criterio de los límites de funcionamiento, se selecciona todos aquellos registros que estén por fuera del rango normal de operación.

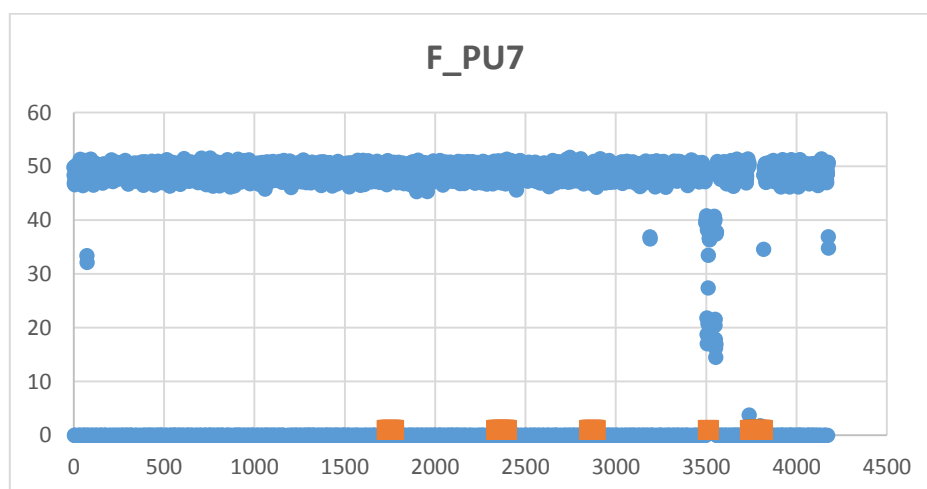


Figura 111. Caudal de la bomba 7 dataset 2

6.2.4 ANALISIS DEL CAUDAL BOMBEADO EN EL SECTOR 4

Como se recuerda, el Sector 4 posee dos bombas en la estación de bombeo, de las cuales solo una está en funcionamiento, la bomba PU8. Como se observa en la Figura 112, donde evidentemente existe una tendencia definida y con una variación mínima del caudal durante todo el tiempo de registro. Con esta disposición, se calcula el valor máximo y mínimo de la franja de registros de tal manera que sirva para la evaluación del criterio.

Así mismo, el dataset 2 muestra ciertas irregularidades en cuanto a la tendencia esperada, por ejemplo en la **Figura 113** entre la hora 2000 y 3000, dos grupos de registros que se alejan en la parte superior y que coinciden con los periodos de datos anómalos del sistema. De este modo se ha de aplicar el criterio a partir del cual todos los valores por fuera de los límites encontrados en el dataset 1 serán etiquetados como sospechosos.

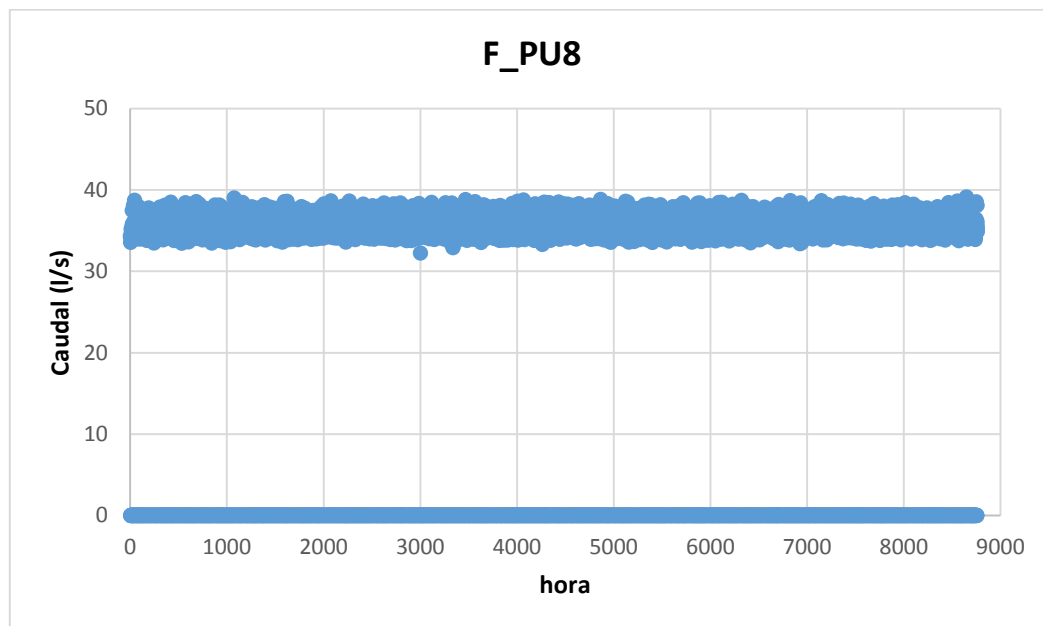


Figura 112. Caudal de la bomba 8 dataset 1

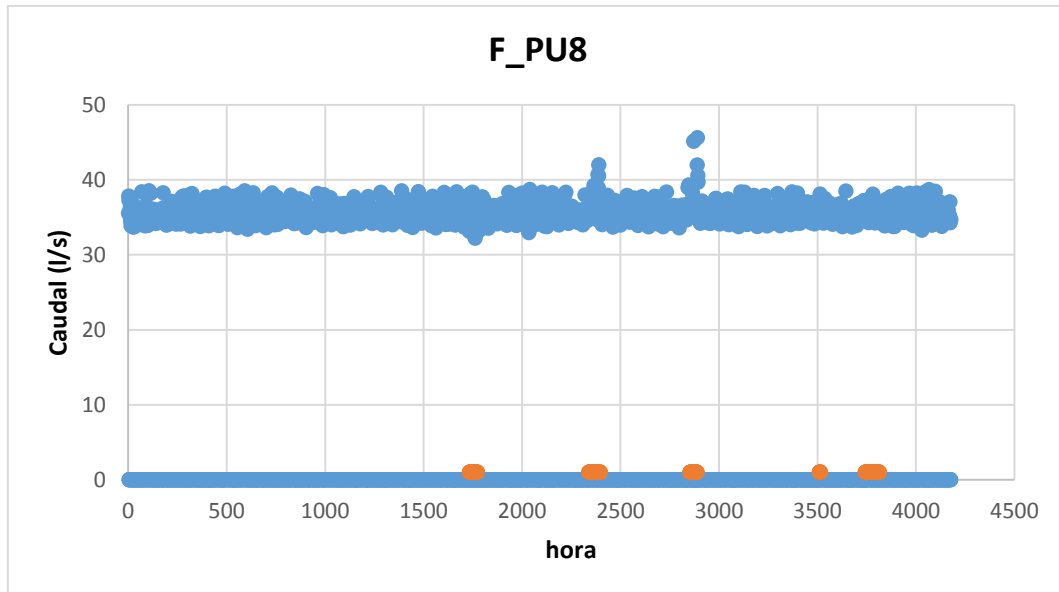


Figura 113. Caudal de la bomba 8 dataset 2.

6.2.5 ANALISIS DEL CAUDAL BOMBEADO EN EL SECTOR 5

Por último, el sector 5 que posee dos bombas en su estación de bombeo: PU10 y PU11, solo tiene en funcionamiento la primera, de modo que se aplicara el criterio solo a esta bomba. En el dataset 1 se ha encontrado una tendencia definida a excepción de un par de valores en la parte inferior como se observa en la **Figura 114**. De esta manera, se tendrán en cuenta estos valores dispersos para el cálculo de los límites de operación de bombeo y así determinar el comportamiento correcto de la bomba PU10.

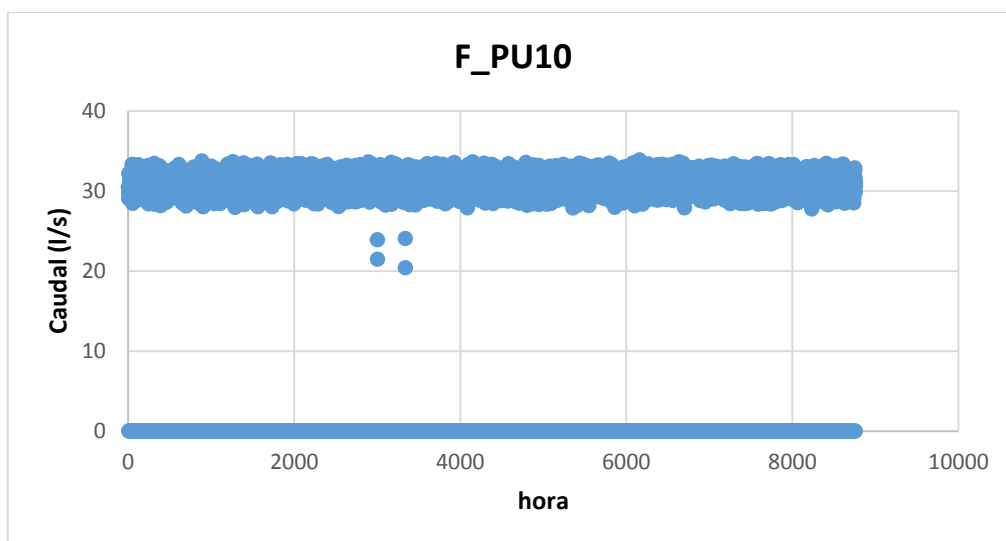


Figura 114. Caudal de la bomba 10 dataset 1

En el dataset 2, por lo contrario, se ha encontrado tres grupos de registros sospechosos, que evidentemente coinciden con los periodos de fallo del sistema como se ve en la Figura 115. De esta forma se prevé tener muchos aciertos a través de la evaluación del criterio para este conjunto de datos, dada la indudable disposición de los valores durante el tiempo de registro.

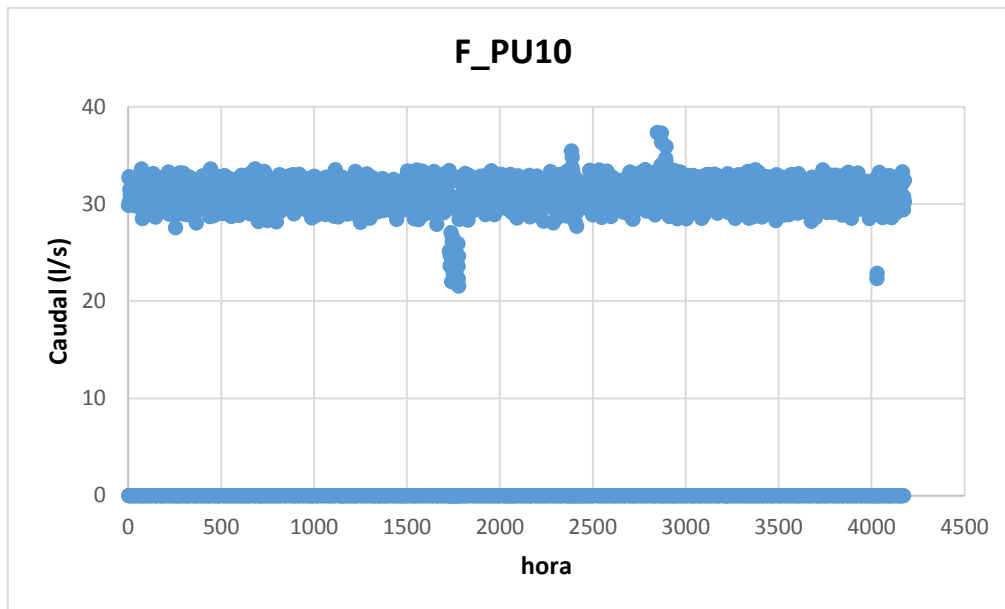


Figura 115. Caudal de la bomba 10 dataset 2

Como se ha observado durante todos los sectores, es claro que la propuesta del análisis del caudal bombeado ha de presentar a simple vista muy buenos resultados, por lo cual es posible intuir un aumento de las detecciones y en consecuencia un aumento de la efectividad del modelo.

6.3 CAUDAL DE LA VALVULA

A parte del análisis de las presiones correspondientes a la válvula, también se planteó un análisis al caudal de la misma, donde se pueda determinar al igual que en las bombas, que tipo de tendencia posee y como se podría luego contrastar con el dataset 2 para obtener algún resultado que nos lleve a la detección de anomalías en la red.

Para la elaboración del análisis, primeramente, se ha determinado el funcionamiento correcto a partir del dataset 1 como se observa en la **Figura 116**, donde se ve un funcionamiento bastante constante en el tiempo y en consecuencia, con marcados límites de operación. La franja inferior sobre el valor cero, nos indica que la válvula no siempre está en funcionamiento, es decir, que existen momentos en los cuales no hay paso de caudal por la misma.

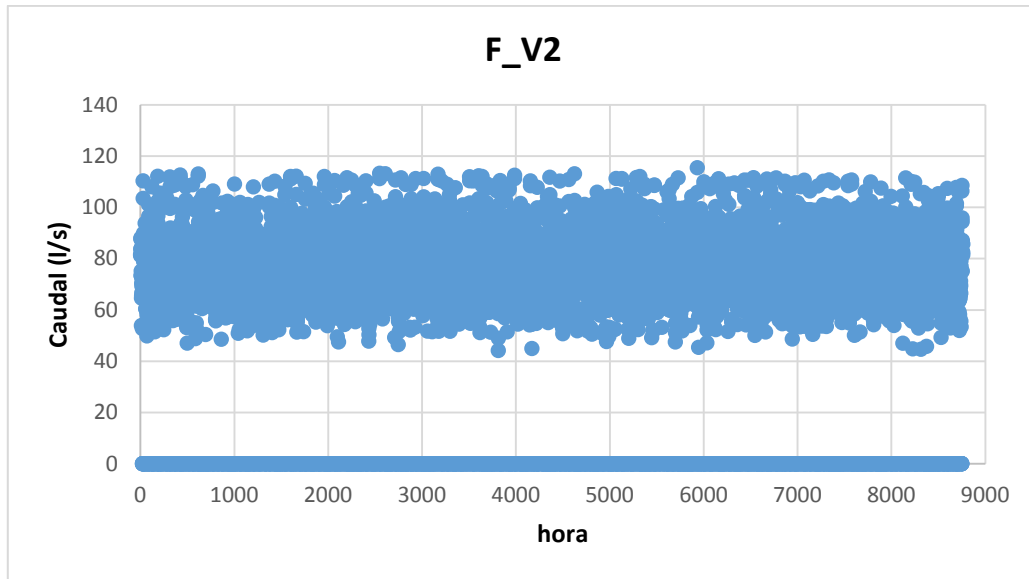


Figura 116. Caudal de la Válvula dataset 1

No obstante, eso no pasa con el dataset 2 como se ve en la Figura 117, donde se nota mucha más variabilidad en los registros y algunas nubes de puntos sospechosas sobre el límite superior que coinciden perfectamente con periodos de fallos en el sistema. Por ejemplo, entre la hora 2000 y 2500, se observan registros que se alejan de la tendencia general, evidenciando que algo sucede en esos instantes y ajustándose precisamente al segundo periodo de fallos.

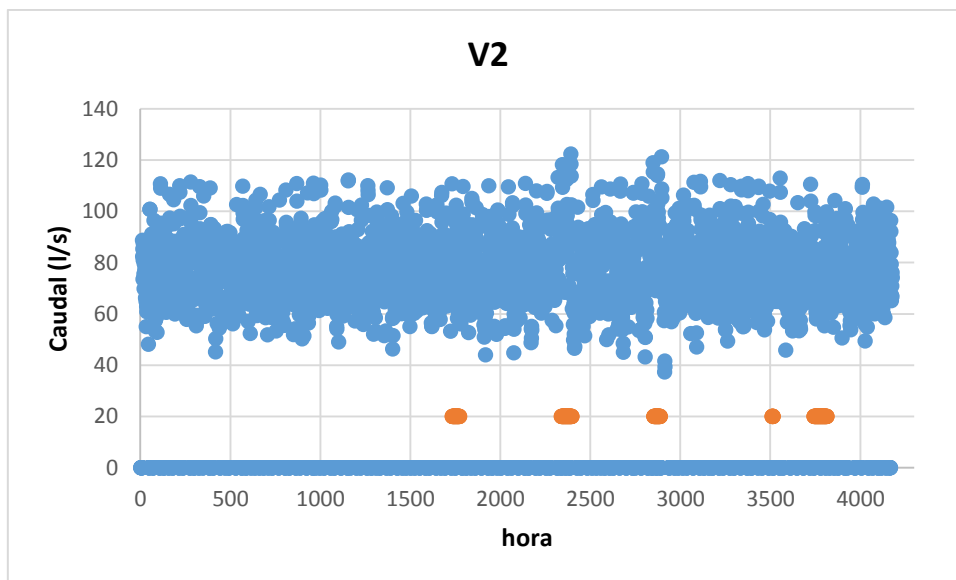


Figura 117. Caudal de la Válvula dataset 2

De esta manera, los valores que no se ajusten a los límites calculados a partir del dataset 1, serán etiquetados como sospechosos, de manera que según lo que se ha podido observar, este criterio posiblemente ayude a incrementar las detecciones en el segundo y tercer periodo de fallos.

6.4 RESULTADO DE LAS MEJORAS AL MODELO

A simple vista se ha podido apreciar que los planteamientos han estado bien enfocados y que se espera buenos resultados, por eso en este punto se recogerá todos los resultados de las mejoras y como ha ayudado al modelo en general para detectar fallos que anteriormente no se había podido lograr.

En primer lugar, se planteó el análisis de las presiones a la entrada y salida, tanto en las estaciones de bombeo como de la válvula reguladora. Luego de haber aplicado el criterio de evaluación correspondiente se ha obtenido los siguientes resultados:

DESCRIPCION	PRESIONES					
	EB1	EB2	EB3	EB4	EB5	V2
SOSPECHOSOS	9	34	9	44	76	20
ACIERTOS	9	5	2	20	39	12
EFFECTIVIDAD	100%	15%	20%	45%	51.3%	60%

Tabla 23. Resultados de la detección por Presiones

Observando los resultados del análisis por presiones, encontramos un criterio con una eficiencia media, donde en el caso de la estación de bombeo 1 se consiguió obtener el 100% de efectividad y en el caso de la válvula que no se había planteado como punto de evaluación, aportó 12 aciertos de 20 sospechas detectadas, por lo cual se obtuvo un resultado bueno. Por otra parte, en las estaciones de bombeo 2 y 3 se encontró las medidas más bajas de efectividad, aunque en el caso de la última, sus valores sospechosos fueron de tan solo 9 instantes.

Como se había planteado anteriormente, todos estos valores de detección serán sumados al valor original que se tenía en un principio en las estaciones de bombeo, cuando se analizó los rangos de funcionamiento y sus curvas características. De esta manera la mejora del criterio queda representada en la siguiente figura:

DESCRIPCION	BOMBAS							
	PU1	PU2	PU4	PU6	PU7	PU8	PU10	PU11
SOSPECHOSOS	7	22	34	90	23	47	76	70
ACIERTOS	7	17	5	61	11	21	37	35
EFFECTIVIDAD	100%	77.27%	15%	68%	25.5%	44.68%	47%	50%

Tabla 24. Resultados finales de la detección por bombeo

Con estos resultados, notamos evidentemente un incremento en cuanto a las sospechas y en consecuencia a los aciertos en las estaciones de bombeo. Por ejemplo, en la bomba PU1 se mantuvo el mismo porcentaje de efectividad, mientras que la bomba PU2 que antes tenía pocas detecciones, ha incrementado ese valor y también su efectividad. Por otro lado, la bomba PU4 que no tenía ninguna detección, ahora ha detectado 5 instantes de fallo, aunque con poca efectividad. También la bomba PU7 ha mantenido el mismo porcentaje de efectividad en comparación a los resultados obtenidos antes de implementar la mejoría, de la misma forma, la bomba PU8 también mantiene su valor de efectividad, no obstante, ha logrado detectar más fallos en el sistema. Por último, para la bomba PU10 y PU11 los valores de efectividad han disminuido, sin embargo, se ha detectado muchos más fallos que al inicio. A continuación, se describe el progreso del modelo en cuanto los aciertos antes y después de la mejoría.

DESCRIPCION	ACIERTOS							
	PU1	PU2	PU4	PU6	PU7	PU8	PU10	PU11
ANTES	7	8	-	59	10	5	30	24
MEJORIA	7	17	5	61	11	21	37	35
TOTAL	0	+7	+5	+2	+1	+16	+7	+11

Tabla 25. Mejorías en las estaciones de bombeo.

Es evidente que las mejorías han dado buenos resultados, dado que en la mayoría de las estaciones de bombeo se ha logrado aumentar los aciertos y consiguiendo así identificar muchas más anomalías que aún no se habían detectado.

Luego de mostrar los números de las mejorías respecto al análisis de las presiones y la válvula, también es necesario saber los resultados de las mejorías respecto al análisis del caudal bombeado. La **Tabla 26** recoge todos los resultados después del análisis del flujo.

DESCRIPCION	CAUDAL								
	PU1	PU2	PU4	PU6	PU7	PU8	PU10	PU11	V2
SOSPECHOSOS	14	7	1	87	4	10	37	70	15
ACIERTOS	10	7	1	59	3	7	27	35	8
EFFECTIVIDAD	71.4%	100%	100%	67.8%	75%	70%	73%	50%	53.3%

Tabla 26. Resultados de la detección por Caudal.

Es notoria la alta efectividad en los resultados, teniendo dos valores del 100% como en el caso de la bomba PU2 y PU4. Además, el valor más bajo encontrado es el de la válvula, con una efectividad del 53.3%; pero en general los resultados encontrados son muy buenos y se espera haber detectado todos los instantes que faltaban a través de este nuevo enfoque.

Como se había comentado en la elaboración del nuevo criterio del caudal en las bombas y válvula, se ha determinado que todos estos resultados se sumen a los resultados encontrados en los patrones de consumo, debido a que el caudal bombeado está directamente relacionado a la cantidad de agua que los usuarios consumen y al llenado y vaciado de los tanques. De este modo se está garantizando una medida clara sobre los consumos reales en la red, de modo que aquellos valores de caudal que sobrepasen los límites, sean considerados potencialmente sospechosos y no como un aumento desmedido del consumo de los abonados. Partiendo de este hecho, a continuación, se muestran los resultados generales del criterio de consumos.

DESCRIPCION	PATRONES Y VALVULA					
	1	2	3	4	5	V2
SOSPECHOSOS	54	1	95	24	63	35
ACIERTOS	23	1	71	11	42	20
EFFECTIVIDAD	43%	100%	74%	45.8%	67%	57.1%

Tabla 27. Resultados finales de la detección por Patrones de consumo.

Estos números a simple vista evidencian un mejor resultado para este criterio, por ejemplo, antes de implementar las mejoras, el patrón 2 y 4 no habían detectado ningún valor sospechoso, sin embargo, luego de implementar este enfoque, las detecciones de los dos han sido muy buenas. Por otro lado, el Patrón 3 ha disminuido su efectividad respecto al resultado mostrado en la **Tabla 18**, donde su efectividad era del 95%; no obstante, a pesar de eso se consiguió elevar el número de aciertos que en principio es lo más importante. De la misma manera, los resultados para el Patrón 5 son mucho mejores, consiguiéndose un aumento en las detecciones y la efectividad. Por último, el análisis de la válvula también ha dado buenos resultados, porque se ha logrado detectar 20 fallos con una efectividad del 57%. Las mejoras en los aciertos se presentan en la **Tabla 28**.

DESCRIPCION	ACIERTOS				
	1	2	3	4	5
ANTES	52	0	61	0	36
MEJORIA	23	1	71	11	42
TOTAL	+1	+1	+10	+11	+6

Tabla 28. Mejoría en los Patrones de Consumo

Obtenidos los resultados finales de cada uno de los criterios, es esencial conocer los resultados generales y las mejoras en la Matriz de resultados. Por ejemplo, en la **Tabla 29** se observa la importancia de las mejoras dentro de la evolución del modelo, donde efectivamente todos los periodos han sido detectados en su totalidad, aumentando la efectividad total del 72% al 100%. Esto demuestra que las mejoras han sido determinantes dentro de todo el desarrollo final del modelo de detección.

PERIODO DE FALLO	DURACIÓN (h)	DETECCIONES (h)	% DETECCION
1	42	42	100
2	60	60	100
3	37	37	100
4	7	7	100
5	73	73	100

Tabla 29. Números finales de detección.

En cuanto a la matriz de resultados, como se había mostrado la **Tabla 29**, indica una detección absoluta de los periodos de fallo, de manera que se detectó 219 instantes y en consecuencia se logró disminuir el valor de los falsos negativos a cero. Por otro lado, la implementación de esta mejora elevó los falsos positivos a 259, por lo cual como se mencionó al principio de este trabajo, lo primordial era detectar completamente las anomalías del sistema, aunque aumenten los valores de este tipo.

MATRIZ DE RESULTADOS		Estado Actual	
		FALLO (positivo)	SEGURO (Negativo)
Estado Predicho	FALLO (positivo)	219	259
	SEGURO (negativo)	0	3699
TOTAL HORAS		4177	

Tabla 30. Matriz final de resultados del Modelo de detección.

Para finalizar, es importante mencionar que conforme el número de detecciones provenientes de los criterios aumenten en una hora determinada, la probabilidad de que estos sean efectivamente fallos aumenta también como se observa en la **Figura 118**. Por ejemplo, cuando tengo detecciones de un solo criterio el porcentaje de que ese valor sea efectivamente un fallo es del 19%, de esta manera, cuando hay dos criterios que me emiten una señal de alarma la probabilidad aumenta hasta el 69%, creciendo de una manera muy súbita. Luego, cuando son tres los criterios que emiten alarmas el porcentaje disminuye un poco hasta el 59% y de ahí en adelante aumenta hasta llegar al 100% cuando se trata de valores de 6 en adelante.

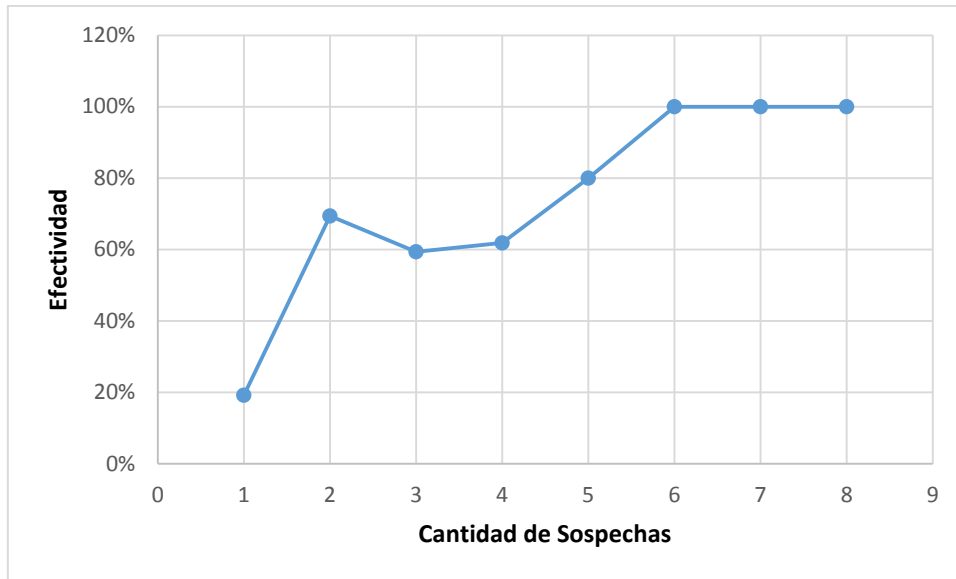


Figura 118. Diagrama de efectividad de las sospechas.

Estos valores demuestran una coherencia dentro del comportamiento general de las detecciones, puesto que es lógico esperar que cuantas más alarmas se tenga en una hora específica, es más la probabilidad de que algo realmente le esté ocurriendo a la red.

7. CONCLUSIONES.

La detección de anomalías ciber-físicas es un campo bastante amplio, donde depende no solo del modelo de detección como tal, sino de un conjunto de elementos que deben estar funcionando en armonía para brindar una seguridad real en las redes de abastecimiento, partiendo en principio de la obtención de información necesaria para poder analizar el sistema, hasta la protección de la infraestructura del sistema.

Por otra parte, para la elaboración de este modelo de fiabilidad aplicado a la red C-Town, se planteó diferentes frentes de trabajo, resultando algunos no tan convenientes, como el análisis de los controles de las bombas, que al final no produjo los resultados esperados; no así con los consumos, que a pesar de no contar con la información del comportamiento de los abonados, produjo muy buenos resultados. Los demás criterios también fueron determinantes para la detección de muchos instantes de fallos durante el desarrollo del modelo, por lo cual, en general la implementación de la mayoría de los criterios fue acertada.

En este mismo sentido, existieron variables que en principio no se consideraron para la elaboración del modelo, por ejemplo, el funcionamiento de la válvula y el análisis individual del caudal de las bombas. Sin embargo, durante toda la elaboración del modelo se usó toda la información suministrada por la red.

También es importante mencionar que el primer análisis que se hizo, produjo resultados relativamente buenos, teniendo una efectividad de hasta el 72% en la detección de los fallos. No obstante, como el objetivo era detectar todos los instantes de anomalías, fue necesario, posterior a los resultados, crear nuevos frentes de detección, que ayuden a complementar el modelo, de tal forma que al final se obtenga una efectividad del 100%. Esto afortunadamente se consiguió y se logró mejorar en muchos de los criterios, la cantidad de detecciones y en otros también su efectividad.

De igual forma, el desarrollo de este modelo ha llevado a la interpretación de muchos de los fenómenos que son muy importantes para la comprensión de la red, por ejemplo, durante la elaboración de los patrones de consumo, se logró entender el comportamiento de los abonados, que aunque al principio se pensaba que podía ser el mismo para toda la red, varía mucho de sector a sector. Por otro lado, con el análisis de las estaciones de bombeo, también se logró determinar las curvas características correspondientes, ayudando así al conocimiento general del sistema.



8. DESARROLLOS FUTUROS.

Para la obtención de los mejores resultados al final del modelo, fue necesario implementar mejoras basadas en el desarrollo de nuevos criterios de evaluación, de modo que después de realizarlas se obtuvo el total de las detecciones para los periodos de fallos. No obstante, existen otros valores dentro de la Matriz de resultados como los falsos positivos, que a futuro pueden ser minimizados y, en consecuencia, abrir un nuevo enfoque de mejoras en el cual se consiga no solamente una efectividad del 100% en las detecciones, sino que el modelo solo alerte cuando realmente ocurra una anomalía en el sistema.

Sería interesante igualmente, el desarrollo de estas estrategias en un programa de simulación de redes de abastecimiento como por ejemplo EPANET, para simular los efectos de los ataques cibernéticos maliciosos y comprobar el alcance al que pueden llegar. Esto también sería importante desde el punto de vista de la prevención de la red, porque se podrían simular múltiples escenarios de fallos y con esto generar las condiciones más desfavorables a las que la red en un momento dado pueda enfrentarse. (Pedro Iglesias, 2012)

Por último, con la propuesta de BATADAL como desafío de diseño, se ha conseguido muy buenos resultados, de modo que sería interesante seguir participando de estos eventos que el Instituto de Recursos Hídricos y Medio Ambiente ha patrocinado durante varios años y así seguir avanzando en la seguridad de las redes de abastecimiento de agua. (Salomon, 2012)



9. BIBLIOGRAFÍA

Institute, E. & (2017). *BATADAL. WORLD ENVIRONMENTAL & WATER RESOURCES CONGRESS*. Sacramento, California.

Marchi, A. (2014). *Battle of the Water Networks II*. Reston, Virginia.

McCaffrey, S. C. (2009). *Convencion de las Naciones Unidas sobre el derecho de los usos de agua Internacionales*. California.

Mundo, E. (2015, 06 17). ¿Qué son los 'Cyber Physical Systems'?

MUNDO, E. (17 de 05 de 2016). España, tercer país del mundo con más ciberataques. pág. 1.

Nils Ole Tippenhauer, A. O. (2016). Simulation of Cyber-Physical Attacks on Water Distribution Systems with EPANET. *Proceedings of the Singapore Cyber-Security Conference (SG-CRC) 2016* (p. 9). Singapore: ResearchGate.

Oliker, N. (2015). *Network hydraulics inclusion in water quality event detection using multiple sensor stations data*. Haifa, Israel: ScienceDirect.

País, E. (2017, 05 16). Europol teme aumento de ciberataques .

Pedro Iglesias, J. S. (2012). *Analisis y diseño de redes de distribucion de agua mediante EPANET 2*. Valencia.

Powers, D. (2011). *EVALUATION: FROM PRECISION, RECALL AND F-MEASURE TO ROC, INFORMEDNESS, MARKEDNESS & CORRELATION*. Flinders, Australia: BioInfo Publications.

Salomon, E. (2012). The Battle of the Water Networks. *Water Distribution Systems Analysis Conference 2012*, (p. 13). Adelaide.

