



UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA ELÉCTRICA

DETECCIÓN DE SUPERNOVAS UTILIZANDO FILTRO DE KALMAN DE MÁXIMA
CORRENTOPÍA Y ENTROPÍA DE CURVAS

MEMORIA PARA OPTAR AL TÍTULO DE
INGENIERO CIVIL ELÉCTRICO

PABLO ANDRÉS MONTERO SALVATIERRA

PROFESOR GUÍA:
PABLO ESTÉVEZ VALENCIA

PROFESOR CO-GUÍA:
FRANCISCO FÖRSTER BURÓN

MIEMBROS DE LA COMISIÓN:
CLAUDIO PÉREZ FLORES
MARCOS ORCHARD CONCHA

SANTIAGO DE CHILE
2020

RESUMEN DE LA MEMORIA PARA OPTAR
AL TÍTULO DE INGENIERO CIVIL ELÉCTRICO
POR: PABLO ANDRÉS MONTERO SALVATIERRA
FECHA: 2020
PROF. GUÍA: PABLO ESTÉVEZ VALENCIA

DETECCIÓN DE SUPERNOVAS UTILIZANDO FILTRO DE KALMAN DE MÁXIMA CORRENTROPÍA Y ENTROPÍA DE CURVAS

Debido al rápido avance de la tecnología, la instrumentación astronómica se ha visto afectada por la construcción de telescopios y radiotelescopios capaces de generar masivas cantidades de información. Esto provoca que los actuales métodos de detección de objetos astronómicos no puedan ser utilizados por su gran costo computacional y alto porcentaje de error. En particular, los métodos relacionados con la detección de supernovas generan un número importante de falsos positivos y negativos, lo que fuerza a la intervención del astrónomo en el proceso. Es por lo anterior, que la demanda de nuevas técnicas capaces de automatizar el proceso de detección son cada vez más necesarias y solicitadas.

En el trabajo de título de Pablo Huentelmu y en el de Paloma Pérez se trabajó en un programa capaz de generar alertas de supernovas a través del uso de Filtros de Kalman en tres versiones: clásico, de máxima correntropía y *unscented*, utilizando como base de datos las series de imágenes obtenidas de las campañas de HiTS 2014A y 2015A. En esta memoria se busca continuar estos trabajos con el objetivo de mejorar los resultados en cuanto a cantidad de supernovas detectadas entre las confirmadas por HiTS, así como cuántos aciertos sostiene y qué tan temprano son detectadas. Además se estudia la posibilidad de simplificar el modelo original.

Para aumentar la cantidad de detecciones se relajaron condiciones utilizadas por la rutina anterior, lo cual permitió detectar más supernovas en comparación al modelo anterior, a un costo de aumentar la cantidad de falsos positivos. Estos últimos fueron filtrados a través de nuevas técnicas como la entropía de curvas de los candidatos y el ajuste paramétrico del Filtro de Kalman de máxima correntropía. Además, el relajamiento aumentó la cantidad de veces que las supernovas eran detectadas y permitió emitir alertas de detección incluso más temprano que HiTS. Estas ventajas se mantuvieron a pesar de la implementación de las técnicas de filtrado ya mencionadas. Finalmente, para simplificar el programa se realizó una poda de las condiciones, es decir, se eliminaron las que no alteraban los resultados finales significativamente, logrando eliminar un 43.5 % de éstas.

A mi abuela

Agradecimientos

Quiero partir agradeciendo a mi madre por todo lo que me ha entregado a lo largo de mi vida, por enseñarme a siempre ver el lado positivo de las cosas. Gracias a sus enseñanzas, apoyo y constante compañía, fui capaz de llegar hasta este punto. Agradezco a mi padrino por formar parte de mi desarrollo personal, del cual he aprendido bastante, por siempre estar presente cuando lo necesité y por todo el ánimo entregado. Agradezco a mi madrina por haberme apoyado, aunque sea a la distancia y por haber sido parte de mi formación como persona. Además, agradezco a toda mi familia por toda la fe y cariño que me han entregado durante todos estos años. Una mención especial a mi abuela, quien fue uno de mis pilares fundamentales a lo largo de gran parte de mi vida, y que a pesar de que ya no está conmigo, aún siento su apoyo.

Agradezco a la Dani, mi polola, por acompañarme durante toda la etapa universitaria. Gracias a ti he logrado superar todos los obstáculos que se han presentado a lo largo de este largo periodo. Gracias por tu alegría y eterna compañía.

Agradezco enormemente a mi profesor guía Pablo Estévez por su paciencia, dedicación y constante acompañamiento. Agradezco a mi profesor co guía Francisco Förster por sus consejos y ayuda.

Agradezco a todos mis amigos del colegio, universidad y de la vida por las risas, los buenos momentos, por siempre estar ahí y la gran cantidad de experiencias compartidas. Agradezco la enorme fe que tienen en mi y su constante apoyo. Además, aprovecho de agradecer a mis gatos por acompañarme en mis largas noches de estudio.

Finalmente quiero agradecer a National Laboratory for High Performance Computing por apoyarme con su infraestructura. Powered@NLHPC: Esta tesis fue parcialmente apoyada por la infraestructura de supercómputo del NLHPC (ECM-02). Además, quiero agradecer a Fondecyt 1171678 por su apoyo.

Tabla de Contenido

1. Introducción	1
1.1. Descripción del problema	2
1.2. Objetivos	2
1.2.1. Objetivo General	2
1.2.2. Objetivos Específicos	2
1.3. Estructura de la Memoria	3
2. Antecedentes	4
2.1. Conceptos Astronómicos	4
2.1.1. Supernovas	4
2.1.2. Magnitud aparente	5
2.2. High Cadence Transient Survey	7
2.3. Filtro de Kalman	8
2.4. Filtro de Kalman de máxima correntropía	10
2.4.1. Correntropía	10
2.4.2. Criterio de máxima correntropía (MCC)	11
2.4.3. Filtro de Kalman de máxima correntropía	12
2.5. Programa	15
2.5.1. Entradas	15
2.5.2. Etapas de filtrado	16
2.5.3. Resultados que se obtienen	16
2.6. Laboratorio Nacional de Computación de Alto Desempeño (NLHPC)	17
2.7. Generalización de la entropía para el análisis de curvas	18
3. Metodología	19
3.1. Metodología general	19
3.2. Base de datos	20
3.3. Modelo de trabajos anteriores	20
3.4. <i>Zero points</i>	21
3.5. Relajación de condiciones	22
3.5.1. Variables acumuladoras	22
3.5.2. Umbrales de flujo y velocidad de flujo	23
3.5.3. ASMR	23
3.5.4. Varianza de flujo	24
3.5.5. Modelo relajado	24
3.6. Entropía de curvas	24

3.7.	Filtro	25
3.8.	Poda	27
3.9.	Otras pruebas	27
3.9.1.	Comparación de modelos obtenidos en distintas secciones	27
3.9.2.	Alta cadencia <i>versus</i> baja cadencia	27
3.9.3.	Imágenes filtradas según <i>airmass</i>	28
3.9.4.	Comparación entre el uso de filtros de Kalman de máxima correntropía y básico	28
3.9.5.	Prueba con otra base de datos	28
4.	Resultados y análisis	29
4.1.	Base de datos	29
4.2.	Resultados de trabajos anteriores	30
4.3.	<i>Zero points</i>	30
4.4.	Relajación de condiciones	32
4.4.1.	Variables acumuladoras	32
4.4.2.	Velocidad de flujo estimada	33
4.4.3.	Flujo estimado	35
4.4.4.	ASMR	35
4.4.5.	Varianza de flujo	36
4.4.6.	Modelo relajado	36
4.5.	Entropía de curvas	39
4.5.1.	Umbral de entropía	39
4.5.2.	Cantidad de puntos para calcular entropía	41
4.5.3.	Modelos relajados con umbrales de entropía	41
4.6.	Filtro de Kalman de máxima correntropía	42
4.6.1.	Parámetro sigma del modelo	42
4.6.2.	Ancho de banda de <i>kernel</i> Gaussiano	43
4.6.3.	Modelo final (sin poda)	45
4.7.	Poda	46
4.7.1.	Modelo final podado	47
4.8.	Otras Pruebas	48
4.8.1.	Comparación de modelos obtenidos en distintas secciones	48
4.8.2.	Imágenes filtradas según <i>airmass</i> y alta cadencia <i>vs</i> baja cadencia	53
4.8.3.	Comparación entre el uso de filtros de Kalman de máxima correntropía y básico	55
4.8.4.	Prueba con otra base de datos	56
	Conclusión	61
4.9.	Trabajo futuro	62
	Bibliografía	64
	A. Reglas de rutina original	66
A.1.	Reglas individuales	66
A.2.	Reglas grupales	66
	B. Resultados y análisis	68

B.1. Tablas 68

Índice de Tablas

4.1. Número de verdaderos positivos (TP), falsos negativos (FN) y falsos positivos (FP) detectados usando los filtros de Kalman básico y de máxima correntropía.	30
4.2. Número de verdaderos positivos (TP), falsos negativos (FN) y falsos positivos (FP) detectados, obtenidos al variar el umbral de velocidad de flujo.	35
4.3. Número de verdaderos positivos (TP), falsos negativos (FN) y falsos positivos (FP) detectados, obtenidos al variar el umbral de flujo.	35
4.4. Número de verdaderos positivos (TP), falsos negativos (FN) y falsos positivos (FP) detectados, obtenidos al variar el parámetro ASMR.	36
4.5. Resumen que muestra los parámetros modificados contrastando los valores del modelo original con los del modelo relajado.	36
4.6. Comparación de cantidad de TP, FN y FP obtenidos al utilizar el modelo original y el modelo relajado.	37
4.7. Diferencias de tiempo (en días) con respecto a HiTS de las supernovas detectadas por la rutina original <i>versus</i> el modelo relajado. Si el valor es negativo significa que es detectado antes que la campaña HiTS, si es positivo significa que fue detectado después que la campaña HiTS, si es 0 significa que fue descubierto al mismo tiempo y es x si no fue descubierto por el modelo respectivo.	38
4.8. Número de verdaderos positivos (TP), falsos negativos (FN) y falsos positivos (FP) detectados, obtenidos al variar el umbral de entropía utilizado como base el modelo relajado.	39
4.9. Número de verdaderos positivos (TP), falsos negativos (FN) y falsos positivos (FP) detectados por modelo original y modelos relajados con distintos umbrales de entropía.	41
4.10. Diferencias de tiempo (en días) con respecto a HiTS de las supernovas detectadas por el modelo original <i>versus</i> los modelos relajados con distintos umbrales de entropía. Si el valor es negativo significa que es detectado antes que la campaña HiTS, si es positivo significa que fue detectado después que la campaña HiTS, si es 0 significa que fue descubierto al mismo tiempo y es x si no fue descubierto por el determinado modelo.	42
4.11. Número de verdaderos positivos (TP), falsos negativos (FN) y falsos positivos (FP) detectados, obtenidos al variar el parámetro sigma del modelo matemático. Se utilizó como base el modelo relajado con umbral de entropía 0,6. . .	43
4.12. Número de verdaderos positivos (TP), falsos negativos (FN) y falsos positivos (FP) detectados, obtenidos al variar $\sigma_{correntropia}$, usando como modelo base el modelo relajado con umbral de entropía 0,3.	44

4.13. Diferencias de tiempo (en días) con respecto a HiTS de las supernovas detectadas por el modelo original <i>versus</i> los modelos relajados con umbral de entropía 0,3, σ_{modelo} 0,4 y distintos valores de $\sigma_{correntropia}$. Si el valor es negativo significa que es detectado antes que la campaña HiTS, si es positivo significa que fue detectado después que la campaña HiTS, si es 0 significa que fue descubierto al mismo tiempo y es x si no fue descubierto por el determinado modelo. . .	45
4.14. Resumen de los parámetros y metaparámetros utilizados por el modelo original y modelo final.	46
4.15. Número de verdaderos positivos (TP), falsos negativos (FN) y falsos positivos (FP) detectados por modelo original y modelos finales con distintos números de condiciones eliminadas.	47
4.16. Diferencias de tiempo (en días) con respecto a HiTS de las supernovas detectadas por el modelo original <i>versus</i> modelos finales con distintas cantidades de condiciones eliminadas. Si el valor es negativo significa que es detectado antes que la campaña HiTS, si es positivo significa que fue detectado después que la campaña HiTS, si es 0 significa que fue descubierto al mismo tiempo y es x si no fue descubierto por el determinado modelo.	48
4.17. Comparación de cantidad de TP, FN y FP obtenidos al utilizar los modelos: original, relajado, final y final podado.	49
4.18. Diferencias de tiempo (en días) con respecto a HiTS de las supernovas detectadas por los modelos original, relajado, final y final podado (7 condiciones eliminadas). Si el valor es negativo significa que es detectado antes que la campaña HiTS, si es positivo significa que fue detectado después que la campaña HiTS, si es 0 significa que fue descubierto al mismo tiempo y es x si no fue descubierto por el determinado modelo.	50
4.19. Comparación de resultados al utilizar el modelo final podado considerando 4 casos. En la columna observaciones se tienen dos opciones, el primero, todas que se refiere a que se consideran las observaciones de todo el periodo de observación, mientras que la segunda opción corresponde a alta cadencia , es decir, solo se considerarán las observaciones encontradas en el período de mayor cadencia. Por otro lado, la columna <i>airmass</i> se refiere a si se consideran o no las observaciones filtradas por este umbral.	55
4.20. Comparación de cantidad de TP, FN y FP obtenidos al utilizar el modelo final variando su filtro.	56
4.21. Diferencias de tiempo (en días) con respecto a HiTS de las supernovas detectadas por el modelo final con filtro de Kalman básico (Básico) <i>versus</i> el modelo final con filtro de Kalman de máxima correntropía (MCC). Si el valor es negativo significa que es detectado antes que la campaña HiTS, si es positivo significa que fue detectado después que la campaña HiTS, si es 0 significa que fue descubierto al mismo tiempo y es x si no fue descubierto por el modelo respectivo.	56
4.22. Número de verdaderos positivos (TP), falsos negativos (FN) y falsos positivos (FP) detectados de la campaña del 2014A de HiTS utilizando los modelos original, final y final podado.	57

B.1. Número de verdaderos positivos (TP), falsos negativos (FN) y falsos positivos (FP) detectados por el modelo original con el factor γ (Con <i>Zero Points</i>) y sin éste (Sin <i>Zero Points</i>).	68
B.2. Número de verdaderos positivos (TP), falsos negativos (FN) y falsos positivos (FP) detectados, obtenidos al variar los parámetros <i>accum neg flux depth</i> (neg) y <i>accum med flux depth</i> (med).	68
B.3. Número de verdaderos positivos (TP), falsos negativos (FN) y falsos positivos (FP) detectados, obtenidos al variar el parámetro de varianza de flujo.	68
B.4. Número de verdaderos positivos (TP), falsos negativos (FN) y falsos positivos (FP) detectados, variando número de puntos para el cálculo del <i>convex hull</i>	69
B.5. Número de verdaderos positivos (TP), falsos negativos (FN) y falsos positivos (FP) detectados, obtenidos al variar el umbral de entropía. Se utilizó como base el modelo relajado.	69
B.6. Número de verdaderos positivos (TP), falsos negativos (FN) y falsos positivos (FP) detectados, variando $\sigma_{correntropia}$ usando como base el modelo relajado.	69
B.7. Número de verdaderos positivos (TP), falsos negativos (FN) y falsos positivos (FP) detectados, realizando poda de condiciones.	69

Índice de Ilustraciones

2.1.	Representación de estructura de capas de una estrella antes de convertirse en supernova. Los elementos que se muestran en la figura son los creados por la estrella a lo largo de su vida. Imagen publicada por R. J. Hall en <i>WikiMedia Commons</i> , 15 de agosto de 2007.	5
2.2.	Imagen proveniente de Dark Energy Cam (DECam), donde se observa a la derecha la explosión de una supernova. https://news.fnal.gov/2013/04/supernovae-light-the-way-to-dark-energy/	6
2.3.	Curvas de luz típicas de supernovas de tipo Ia y tipo II. ©2006 <i>Pearson Education Inc.</i> , publishing as <i>Addison-Wesley The Bizarre Stellar Graveyard</i>	6
2.4.	Diagrama ilustrativo del brillo de una supernova <i>versus</i> el tiempo. En el cuadro blanco se encierra el fenómeno elusivo de <i>Shock Breakout</i> . https://phys.org/news/2016-03-astrophysicists-supernovae-moment-explosion.html	7
2.5.	Imagen real del DECam, cámara de 520 Mp, donde se aprecian los 62 CCDs utilizados. https://www.turismodeestrellas.com/noticias/astronomia/3539-dark-energy-survey-des-el-fabuloso-cartografo-del-cielo.html	8
2.6.	Representación de los 62 CCDs del DECam, donde se observa la existencia de detectores que comienzan con S o N, los cuales son los encargados de obtener las imágenes científicas. http://www.ctio.noao.edu/noao/node/2250	9
2.7.	Representación de funcionamiento de la rutina propuesta por autores anteriores para detectar supernovas. El programa inicia leyendo la serie de imágenes correspondientes a una época en particular, donde se analiza si los píxeles individuales de cada una de estas cumple con el Filtro de píxeles individuales . Los píxeles que superan las 7 condiciones durante 4 épocas consecutivas pasarán a una etapa de Agrupamiento , donde se agruparan los píxeles según su cercanía. A partir de los nuevos grupos, se determinarán sus centros para ser filtrados por el Filtro de grupo de píxeles . Finalmente, si la serie de imágenes y los píxeles centrales cumplen con 9 nuevas reglas se emitirá una alerta.	17
3.1.	Diagrama que representa la metodología utilizada en este trabajo. Se inicializa relajando las condiciones del programa original para obtener nuevos candidatos, luego se implementa una nueva condición y se ajustan parámetros para filtrar los falsos positivos obtenidos del paso anterior, finalmente se procede a podar y a realizar otras pruebas que serán mencionadas más adelante.	20
3.2.	Cadencia de observaciones de campaña 2014A (5 noches).	21
3.3.	Cadencia de observaciones de campaña 2015A (11 noches).	21

3.4.	Curva de espacio de estados de la supernova encontrada en el campo 02 y CCD N5. En azul se observa al curva estimada por el filtro de Kalman de máxima correntropía y en rojo se observa la curva de los valores observados.	25
3.5.	Etapas para obtener el <i>convex hull</i> . (a) Muestra los puntos de la curva en el espacio de estados de una supernova, (b) muestra los vértices de la envolvente convexa y (c) muestra la envolvente convexa.	26
4.1.	Conjunto de gráficos que muestran información referente al píxel en el que se emitió una alerta. De arriba hacia abajo se observa que el primer gráfico muestra el comportamiento del flujo en ADU a través del tiempo expresado en MJD, se muestra el estado observado (rojo), el estado predicho (verde) y estado estimado (azul) con sus respectivos errores. El segundo gráfico muestra la variable de estado de velocidad de flujo en ADU/días considerando su curva predicha (verde) y estimada (azul). El tercer gráfico muestra qué condiciones no fueron satisfechas por el píxel en estudio. Finalmente, en el último gráfico se muestra información sobre la varianza de las componentes predichas y estimadas.	31
4.2.	Gráfico que muestra la curva en el espacio estados del píxel en estudio. La curva estimada está representada en azul mientras que la observada en rojo. Se aprecia además unas líneas rectas discontinuas que corresponden a los umbrales que deben ser superados para que estos puntos sean considerados como candidatos. Cabe señalar que en la leyenda de arriba a la derecha se agrega la información de la entropía.	32
4.3.	Serie de estampillas que muestran la información alrededor del píxel en estudio por época, donde se muestran (de arriba hacia abajo) las imágenes científicas, imágenes de diferencia, PSF, flujo observado, varianza del flujo observado, flujo estimado, velocidad de flujo estimada, grupo de píxeles candidatos alrededor del central, grupos que son descartados, píxeles que son defectuosos (<i>base mask</i>).	33
4.4.	Comparación de curvas de flujo en ADU con y sin el factor de corrección de la supernova encontrada en el campo 02 y CCD N5.	34
4.5.	Comparación de curvas de flujo en ADU con y sin el factor de corrección de la supernova encontrada en el campo 28 y CCD S16.	34
4.6.	Cantidad de detecciones por supernova descubierta en el conjunto de prueba, comparando modelo original con el modelo relajado.	37
4.7.	Curvas de flujo obtenidas de supernovas que no fueron encontradas por el modelo original pero sí por el modelo relajado. (a) curva de flujo observado, (b) curva de flujo estimado.	38
4.8.	Gráfico <i>recall vs precision</i> ante distintos modelos en los que se varía solo el umbral de entropía. Se incluye además los resultados de autores anteriores, los cuales corresponden a los puntos marcados con la palabra “original”.	40
4.9.	Cantidad de detecciones por supernova descubierta en el conjunto de prueba, comparando el modelo original con los modelos relajados con distinto umbral de entropía.	42
4.10.	Curvas de flujo estimadas por el filtro de Kalman de máxima correntropía, predichas y observadas de supernova ubicada en el campo 27, CCD 10, perteneciente al conjunto de entrenamiento. En (a) se observa la curva de flujo utilizando $\sigma_{modelo} = 0,1$, mientras que en (b) se observa esta misma curva, utilizando $\sigma_{modelo} = 0,4$	44

4.11. Cantidad de detecciones por supernova descubierta en el conjunto de prueba. Se compara modelo original con modelos que utilizan distintos umbrales de entropía.	45
4.12. Cantidad de detecciones por supernova descubierta en el conjunto de prueba, comparando modelo original con modelo finales con distintas cantidades de condiciones eliminadas.	48
4.13. Cantidad de detecciones por supernova descubierta en el conjunto de pruebas (<i>test</i>), comparando los modelos: original, relajado, final y final podado. . . .	49
4.14. Estampillas correspondientes a los píxeles candidatos de la supernova del conjunto de entrenamiento ubicada en el campo 07 y CCD N14. En (a) se encuentra la serie correspondiente al modelo original, en (b) se encuentra la serie que corresponde al modelo relajado y en (c) se encuentra la serie correspondiente al modelo final podado.	50
4.15. Ejemplos de curvas de flujo del conjunto de entrenamiento, descubiertas por los modelos original, relajado, final y final podado.	51
4.16. Ejemplos de curvas de flujo del conjunto de prueba, descubiertas solo por los modelos relajado, final y final podado.	51
4.17. Ejemplos de curvas en espacio de estados de supernovas pertenecientes al conjunto de entrenamiento, descubiertas por todos los modelos original, relajado, final y final podado. Además, se aprecian 2 líneas rectas discontinuas de color negro, las cuales corresponden a los umbrales de flujo (200 [ADU]) y de velocidad de flujo (20 [ADU/días]).	52
4.18. Ejemplos de curvas en sus espacio de estados de supernovas del conjunto de prueba, descubiertas solo por los modelos relajado, final y final podado. Además, se aprecian 2 líneas rectas discontinuas de color negro, las cuales corresponden a los umbrales de flujo (200 [ADU]) y de velocidad de flujo (20 [ADU/días]).	53
4.19. Imágenes correspondiente a un falso positivo encontrado por el modelo final, el cual correspondería a un candidato a estrella variable.	54
4.20. Imágenes correspondiente a un falso positivo encontrado por el modelo final, el cual se produce por la existencia de un crecimiento aparente en su flujo. . .	58
4.21. Imágenes correspondiente a un falso negativo, el cual se produce por una observación tardía del objeto.	59
4.22. Cantidad de detecciones por supernova descubierta en el conjunto de prueba, comparando el modelo original con los modelos finales con distinto filtro. . .	60

Capítulo 1

Introducción

La astronomía, una ciencia que depende de las observaciones y mediciones de objetos del universo, se ha visto fuertemente afectada por el surgimiento de nuevos telescopios que poseen cámaras más potentes, que permiten recolectar masivas cantidades de información. Algunos ejemplos de estos instrumentos son: el *Panoramic Survey Telescope and Rapid Response System* (PanSTARRS) [1], el *Visible and Infrared Survey Telescope for Astronomy* (VISTA) [2], el *Dark Energy Camera* (DECam) [4], entre otros. Por otro lado, existen proyectos de nuevos instrumentos astronómicos que generarán aún más información de la que se ha recopilado hasta ahora, como es el caso del *Large Synoptic Survey Telescope* (LSST) [3], telescopio de alto campo de visión, ubicado en Chile, que posee el potencial para registrar el cielo visible cada 3 noches generando así más de 30 TB de información y alrededor de diez millones de alertas de objetos astronómicos por noche, tales como supernovas, cometas, asteroides, estrellas variables, entre otros.

Dado lo anterior, las técnicas actuales de detección de objetos astronómicos se han vuelto obsoletas por su gran costo computacional y la necesidad de que astrónomos deban revisar visualmente miles de imágenes con el fin de determinar si corresponde a un elemento de interés o a uno de una o más categorías. Esto ha provocado un aumento en la demanda de nuevas técnicas que sean capaces de extraer información de forma rápida, eficiente y robusta, ante problemas inherentes a las mediciones, con el objetivo de disminuir el número de imágenes que los expertos deban revisar para corroborar la categoría a la que corresponde el objeto o, en el mejor de los casos, automatizar este proceso.

Con el fin de detectar objetos astronómicos transitorios cuyas escalas de tiempo se encuentran en el rango de horas a días, se han desarrollado una serie de sondeos como es lo es el caso del *High Cadence Transient Survey* (HiTS). Gracias a éste, se obtuvo una secuencia de imágenes provenientes del *Dark Energy Camera* (DECam) con las que se trabajó en la detección y reporte de candidatos de supernovas. Para realizar lo anterior, se identificaron los elementos de interés y acotaron su estudio a una estampilla de dimensión de 21 x 21 alrededor del objeto, los cuales fueron procesados y posteriormente clasificados a través del algoritmo *Random Forests* [5].

La cantidad de alertas de detección de supernovas dependerá tanto de la calidad de los

datos como de los métodos de filtrado y procesamiento de la información. Es por esto, que los autores Pablo Huentelenu [6] y Paloma Pérez [7] utilizaron filtros de Kalman en su versión básica, de máxima correntropía y *Unscented* para desarrollar una rutina capaz de detectar supernovas.

En esta memoria, se trabaja sobre la rutina creada por los autores mencionados anteriormente y se propone estudiar un método basado en el cálculo de la entropía de curvas con el objetivo de introducir nuevas formas para descartar objetos que no son supernovas como lo son los asteroides, rayos cósmicos, entre otros, los cuales comúnmente generan alertas falsas en los sistemas de detecciones actuales.

1.1. Descripción del problema

La presente memoria busca continuar con los trabajos de título anteriores de Pablo Huentelenu [6] y Paloma Pérez [7], quienes desarrollaron un programa capaz de detectar supernovas basado en distintas versiones del filtro de Kalman: filtro de Kalman básico, filtro de Kalman de máxima correntropía y filtro de Kalman *Unscented* con el fin de estimar las variables de estado de flujo y velocidad de flujo, los cuales utilizan además una serie de reglas adicionales para corroborar si el objeto en estudio corresponde o no a una supernova.

Si bien es cierto, el algoritmo desarrollado por los autores ya mencionados es capaz de detectar supernovas con un bajo número de falsas alarmas, se busca determinar si es posible encontrar aún más a través del relajamiento de reglas, ajuste paramétrico y la implementación del cálculo de entropía de curvas. Además, se explorará la posibilidad de detectar supernovas más tempranamente que el modelo original.

1.2. Objetivos

1.2.1. Objetivo General

Detectar supernovas utilizando filtro de Kalman clásico y de máxima correntropía, usando la entropía de curvas para disminuir los errores de tipo I y tipo II.

1.2.2. Objetivos Específicos

- Desarrollar métodos capaces de distinguir entre supernovas (verdaderos positivos) y otros objetos (falsos positivos) basado en el cálculo de la entropía de la curva en el espacio de estados.
- Optimizar parámetros y condiciones del modelo actual y los relacionados al filtro de Kalman de máxima correntropía.
- Comparar resultados obtenidos en este trabajo con los de autores anteriores y con los del sondeo HiTS, utilizando como criterios la cantidad de detecciones correctas, diferencia de tiempo de alerta (con respecto a HiTS) y la cantidad de veces que cada supernova es detectada.

1.3. Estructura de la Memoria

A continuación se describe la estructura del documento:

- **Capítulo 2** Antecedentes: En este capítulo se presentarán los conceptos básicos necesarios que serán utilizados en el trabajo.
- **Capítulo 3** Metodología: Se describe el procedimiento a realizar para cumplir con los objetivos del trabajo.
- **Capítulo 4** Resultados y análisis: Se muestran y analizan los resultados obtenidos a partir de la metodología propuesta, realizando comparaciones con el modelo de trabajos anteriores.
- **Capítulo 5** Conclusiones: Se presentan las conclusiones del trabajo de memoria, explicando si se logran o no cumplir los objetivos además de proponer trabajo futuro.
- **Bibliografía:** Sección en la que se citan los trabajos utilizados para llevar a cabo la presente memoria.

Capítulo 2

Antecedentes

En este capítulo se abordan los antecedentes claves utilizados en el desarrollo del trabajo. Se inicia con la definición de conceptos astronómicos, para luego describir el sondeo *High Cadence Transient Survey* (HiTS) del cual se obtiene la base de datos con la que se trabajó. Luego, se dan los fundamentos del filtro de Kalman de máxima correntropía y el cálculo de entropía de curvas. Además, se describe el programa con el que se trabaja y el *cluster* de computación donde se desarrollan las pruebas.

2.1. Conceptos Astronómicos

2.1.1. Supernovas

Una estrella, en su ciclo normal, es un cuerpo gaseoso que existe gracias al equilibrio de dos fuerzas que se encuentran en oposición: la gravedad y su presión interna, la cual es producida por fusiones nucleares.

Las supernovas se pueden clasificar, entre otras categorías, en supernovas de tipo Ia y de tipo II en donde la diferencia se encuentra en que las primeras se originan por la explosión de enanas blancas las cuales terminan desapareciendo completamente y las supernovas de tipo II que provienen de estrellas de alta masa.

Las supernovas de tipo II parten como estrellas compuestas principalmente por hidrógeno el cual es transmutado en helio. Una vez que se acaba el hidrógeno, el núcleo de la estrella se contrae y las fusiones nucleares logran transformar el helio en carbono, donde nuevamente el núcleo se contrae y a través de las fusiones nucleares que siguen ocurriendo, se logra transformar el carbono en oxígeno y así sucesivamente hasta que se forman elementos como el hierro o el níquel. La Figura 2.1 representa a las estrellas por capas en donde los elementos más pesados se encuentran en el centro y los más livianos alejados de éste. Es necesario que una estrella posea más de 8 masas solares para que puedan ocurrir todas las reacciones mencionadas anteriormente hasta llegar al hierro.

Cuando la estrella intenta fusionar hierro, la presión interna empieza a disminuir ya que es una reacción endotérmica, es decir, se requiere energía para producir la fusión de este

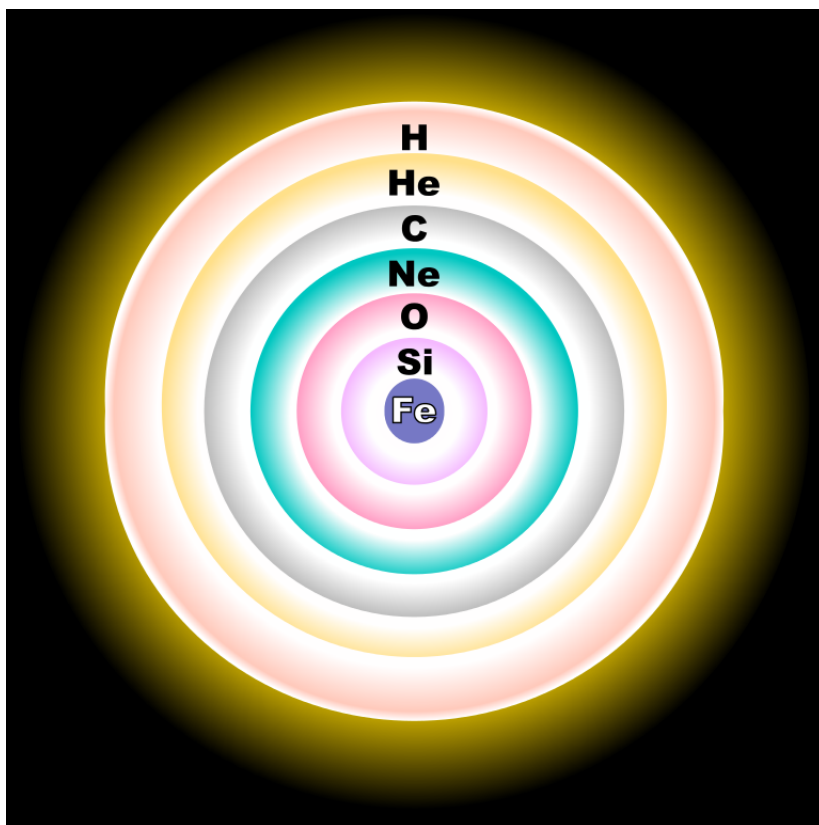


Figura 2.1: Representación de estructura de capas de una estrella antes de convertirse en supernova. Los elementos que se muestran en la figura son los creados por la estrella a lo largo de su vida. Imagen publicada por R. J. Hall en *WikiMedia Commons*, 15 de agosto de 2007.

elemento. Lo anterior provoca que la estrella colapse por su propio peso y como consecuencia de esto las capas más externas de la estrella son expulsadas rápidamente en una gran explosión conocida como supernova. Dicha explosión puede ser observada en la Figura 2.2, donde se observa la misma región del espacio en dos instantes de tiempo diferente, donde en la imagen de la derecha se aprecia un gran aumento de luminosidad en un punto, cual se trata de una supernova.

Una supernova puede ser estudiada a través de su curva de luz, gráfico del brillo de la fuente en función del tiempo, puesto que estas tienen una forma característica, en las cuales se observa un rápido incremento en su luminosidad y/o en su flujo, seguido de un decaimiento. En la Figura 2.3 se observan las curvas de luz de una supernova de tipo Ia y de tipo II en donde una gran diferencia es que la primera presenta un solo decaimiento, mientras que la última posee dos.

2.1.2. Magnitud aparente

La magnitud aparente (m) corresponde a la medida del brillo percibido de una estrella vista desde la Tierra en función del flujo recibido de esta. Su ecuación es:



Figura 2.2: Imagen proveniente de Dark Energy Cam (DECam), donde de observa a la derecha la explosión de una supernova. <https://news.fnal.gov/2013/04/supernovae-light-the-way-to-dark-energy/>.

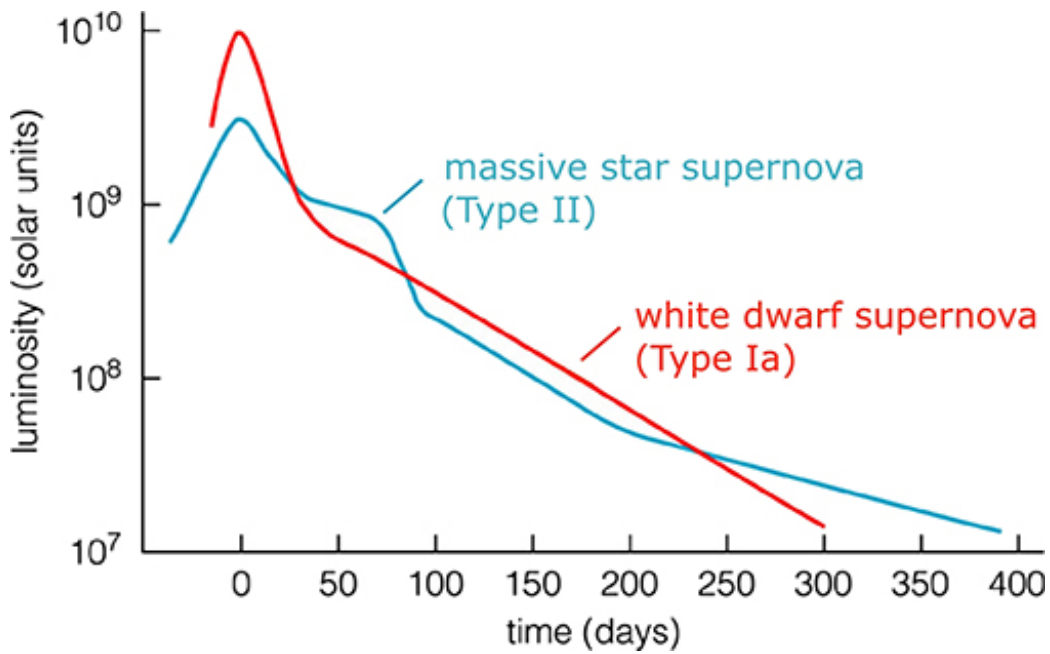


Figura 2.3: Curvas de luz típicas de supernovas de tipo Ia y tipo II. ©2006 Pearson Education Inc., publishing as Addison-Wesley *The Bizarre Stellar Graveyard*.

$$m = -2,5 \log_{10} \frac{F_s}{F_{ref}}, \quad (2.1)$$

donde F_s es la medida del flujo de la fuente y F_{ref} es un flujo de referencia. Además, puede ser reescrita de la forma:

$$m = -2,5 \log_{10}(F_s) + zp, \quad (2.2)$$

donde el valor zp corresponde al denominado *zero point*, el cual es utilizado para corregir distintos fenómenos de la atmósfera como lo son la absorción y la difracción del color.

2.2. High Cadence Transient Survey

El *High Cadence Transient Survey*, desde ahora llamado HiTS, es un *survey* cuyo principal objetivo fue detectar el elusivo fenómeno conocido como *Shock Breakout* (SBO) [5]. Éste es un evento que puede durar desde horas a días y consiste en una onda de choque generada cuando la fuerza de presión interna de la estrella es vencida por la fuerza de gravedad. Una representación de éste se muestra en la Figura 2.4. Pese a los esfuerzos y campañas realizadas no logró ser encontrado. Otro de sus objetivos fue detectar y realizar seguimiento de objetos astronómicos cuyas escalas de tiempo rondaran entre las horas a días, en particular, supernovas.

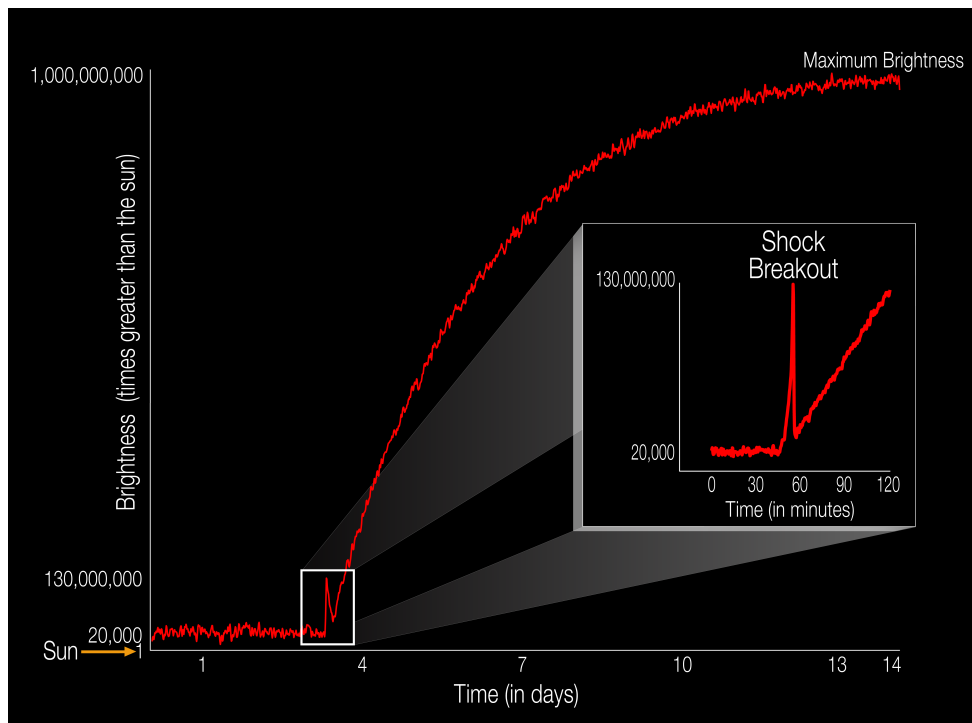


Figura 2.4: Diagrama ilustrativo del brillo de una supernova *versus* el tiempo. En el cuadro blanco se encierra el fenómeno elusivo de *Shock Breakout*. <https://phys.org/news/2016-03-astrophysicists-supernovae-moment-explosion.html>.

Este *survey* utilizó la cámara de energía oscura representada en la Figura 2.5 (DECam por sus siglas en inglés), la cual se encuentra montada en el Telescopio Blanco del Observatorio de Cerro Tololo, región de Coquimbo, Chile.

Esta cámara posee 62 detectores CCD de 2048x4096 píxeles cada uno destinados a la obtención de imágenes científicas y un canal personalizado para la sustracción de imágenes,

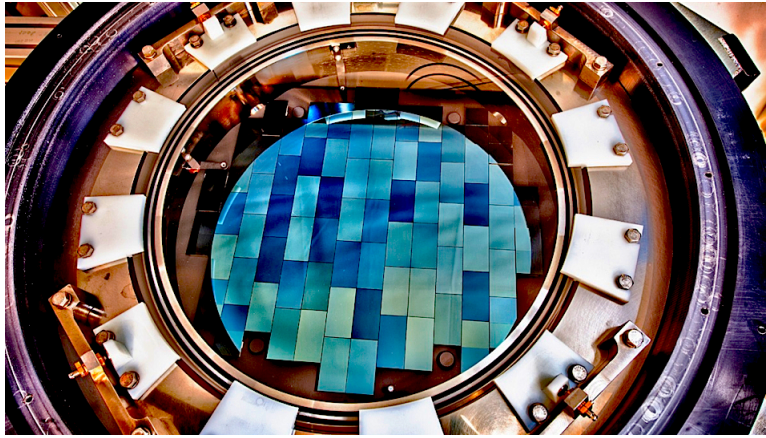


Figura 2.5: Imagen real del DECam, cámara de 520 Mp, donde se aprecian los 62 CCDs utilizados. <https://www.turismodeestrellas.com/noticias/astrologia/3539-dark-energy-survey-des-el-fabuloso-cartografo-del-cielo.html>.

filtrado de candidatos y posterior visualización de candidatos que se ejecutan de forma *online*. En la Figura 2.6 se observa una representación de los CCDs junto a su orientación sobre el cielo.

A través de este *survey* se realizaron 3 campañas comprendidas en los años 2013, 2014 y 2015 entre los cuales se lograron detectar 120 candidatos de supernovas. Mientras que las campañas de los años 2013 y 2014 observaron 40 campos diferentes del cielo, la campaña del 2015 observó 50 campos.

Como cada campo es observado por 62 CCDs, la cantidad de series de imágenes obtenidas por las campañas del 2014 y 2015 respectivamente son: $40 \times 62 = 2480$ y $50 \times 62 = 3100$. Para este trabajo en particular solo se utilizan las series de imágenes en donde fueron detectadas supernovas, reduciendo la cantidad de estas series de 2480 a 32 y de 3100 a 89 (el resto de los datos fueron eliminados).

2.3. Filtro de Kalman

El Filtro de Kalman [8] es un estimador óptimo, que permite determinar los estados ocultos de un sistema lineal utilizando información proveniente de procesos y mediciones contaminadas con ruidos aditivos cuya distribución sea Gaussiana. Para realizar la estimación de los estados, el método inicia con la predicción (también denominado estimación a priori del estado) del estado basándose en el conocimiento previo del sistema. Luego, dicha predicción es comparada con una medición, asociada al estado, en una fase de corrección, la cual representa un promedio ponderado del valor a priori junto al valor medido. Finalmente, lo que se obtiene, es una estimación a través de la combinación de la predicción y la medición basándose en las covarianzas del proceso y del ruido asociado a la medición.

Sea $x(k)$ un vector de largo N , el cual representa el conjunto de variables de estados estimados en el instante k , la predicción es representada por:

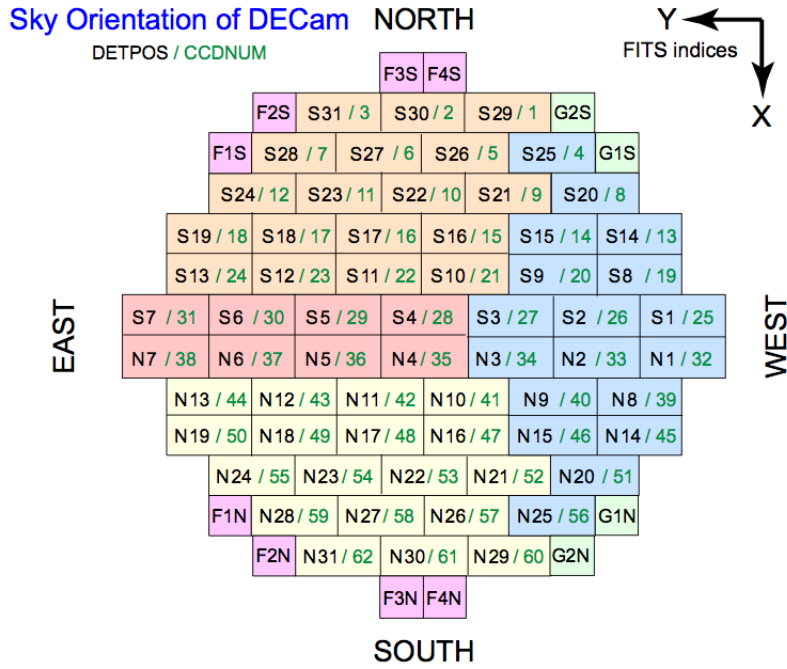


Figura 2.6: Representación de los 62 CCDs del DECam, donde se observa la existencia de detectores que comienzan con S o N, los cuales son los encargados de obtener las imágenes científicas. <http://www.ctio.noao.edu/noao/node/2250>.

$$x(k) = F(k)x(k-1) + B(k)u(k) + q(k), \quad (2.3)$$

donde $q(k)$ representa al ruido aditivo de distribución Gaussiana del proceso con media 0 y matriz de covarianza $Q(k)$. Además, la matriz $F(k)$ representa la matriz de transición de estado, de dimensión $N \times N$ y la matriz $B(k)$ es la matriz de control de entrada, de dimensión $N \times L$, donde L corresponde al largo del vector $u(k)$, la cual representa a una entrada de control. Para el caso de esta memoria, se considerará que $u(k) = 0$

Luego, se tiene las mediciones $y(k)$ de largo M (M puede ser menor a N) que dependen del estado predicho y de la matriz de transformación de estado a medición H , de dimensiones $M \times N$. Al igual que para el caso anterior se supone que esta medición está contaminada por un ruido aditivo de distribución Gaussiana $r(k)$ de media 0 y matriz de covarianza $R(k)$. Lo descrito anteriormente se puede apreciar en la ecuación:

$$y(k) = H(k)x(k) + r(k). \quad (2.4)$$

Este filtro busca minimizar la matriz de covarianza $R(k)$ del ruido de la medición $r(k)$ y la matriz de covarianza $Q(k)$ del error de predicción $q(k)$, para lo cual se utiliza la siguiente función de costo cuadrática (J):

$$\begin{aligned}
J = & \frac{1}{2}(y(k) - H(k)x(k))^T R_k^{-1}(y(k) - H(k)x(k)) \\
& + \frac{1}{2}(x(k) - F(k)x(k-1))^T Q^{-1}(k)(x(k) - F(k)x(k-1)).
\end{aligned} \tag{2.5}$$

Es posible derivar el Filtro de Kalman a partir de la resolución de la Ecuación 2.6.

$$\frac{\partial J}{\partial x(k)} = 0. \tag{2.6}$$

A partir de esta fórmula es posible obtener la mejor estimación del estado, la cual corresponde a:

$$x(k) = F(k)x(k-1) + K(k)H^T(k)(y(k) - H(k)F(k)x(k-1)), \tag{2.7}$$

donde K es la ganancia de Kalman que se puede obtener a partir de la siguiente ecuación:

$$K(k) = (H^T(k)R(k)^{-1}H(k) + Q(k)^{-1})R(k)^{-1}. \tag{2.8}$$

Esta derivación es óptima cuando se asume que los ruidos del proceso y de las mediciones presentan distribuciones Gaussianas. Dado que en la astronomía el ruido de las mediciones normalmente poseen una distribución impulsiva o de cola ancha se hace necesario utilizar otras funciones de costo con las cuales derivar el filtro.

2.4. Filtro de Kalman de máxima correntropía

El filtro de Kalman básico es derivado a partir del error cuadrático medio (MSE, por sus siglas en inglés), y es óptimo bajo supuesto de Gaussianidad y linealidad del sistema. El problema de esto, es que los datos astronómicos presentan distribuciones de cola ancha, lo que provoca que los resultados entregados por el filtro no sean óptimos, además, el MSE es sensible a *outliers*. Por lo anterior, se requiere cambiar la función de costo con la cual se deriva el filtro de Kalman por una capaz de solucionar los problemas mencionados anteriormente, como lo es, la métrica de teoría de la información denominada correntropía. Esta métrica es capaz de entregar robustez a la presencia de *outliers* dado que considera momentos de orden par y superior a 2.

2.4.1. Correntropía

La correntropía [9] es una medida de similitud entre dos variables aleatorias continuas X e Y $\in \mathbb{R}$ que se define matemáticamente como:

$$V(X, Y) = E[\kappa(X, Y)] = \int \kappa(x, y) dF_{XY}(x, y), \tag{2.9}$$

donde κ es un *kernel* Mercer, es decir, es simétrico y semi definido positivo que es invariante a desplazamientos y $F_{XY}(x, y)$ corresponde a la distribución de probabilidad conjunta. Para efectos de este trabajo se empleará la función de *kernel* Gaussiana G_σ la cual se expresa como sigue.

$$\kappa(x, y) = G_\sigma(e) = \exp\left(-\frac{e^2}{2\sigma^2}\right), \quad (2.10)$$

donde $e = x - y$ y $\sigma > 0$ representa el ancho del *kernel*. En situaciones prácticas, la función de distribución conjunta F_{XY} no se conoce, por lo que se procede a estimar la correntropía utilizando un estimador del promedio sobre la muestra a través de la siguiente ecuación,

$$\hat{V}(X, Y) = \frac{1}{N} \sum_{i=1}^N G_\sigma(e_i). \quad (2.11)$$

Análogamente al caso anterior, $e_i = x_i - y_i$ y N es el número de muestras extraídas de F_{XY} . Finalmente, a través del uso de la expansión en series de Taylor, la correntropía puede ser expresada como sigue:

$$V(X, Y) = \sum_{n=0}^{\infty} \frac{(-1)^n}{2^n \sigma^{2n} n!} E[(X - Y)^{2n}]. \quad (2.12)$$

Esta expansión es utilizada para mostrar propiedades importantes de la correntropía como:

- Se consideran todos los momentos pares de la diferencia de sus entradas, incluyendo la información de la correlación normal (que corresponde al caso $n = 1$)
- El parámetro σ controla el énfasis de los momentos de orden mayor en comparación a los de segundo orden.

En otras palabras, el uso de la correntropía permite trabajar con una mayor cantidad de información (momentos pares) en comparación al error cuadrático medio que solo trabaja con el momento de segundo orden. Además, se tiene que si σ posee valores muy altos, la medida se aproxima a la correlación, es por esto y por las razones anteriores que se dice que la correntropía es una generalización de la función de correlación.

2.4.2. Criterio de máxima correntropía (MCC)

A partir de la secuencia de errores $\{e(i)\}_{i=1}^N$ es posible definir una función de costo:

$$J_{MCC} = \frac{1}{N} \sum_{i=1}^N G_\sigma(e(i)). \quad (2.13)$$

Suponiendo que el objetivo es aprender un vector de parámetros W de un modelo adaptativo, donde $x(i)$ corresponde a la salida del modelo e $y(i)$ corresponde a la respuesta deseada, el aprendizaje basado en MCC puede ser formulado a partir de la ecuación propuesta en [10]:

$$\hat{W} = \arg \max_{W \in \Omega} \frac{1}{N} \sum_{i=1}^N G_{\sigma}(e(i)), \quad (2.14)$$

donde \hat{W} es la solución óptima y Ω denota el conjunto de parámetros factibles.

2.4.3. Filtro de Kalman de máxima correntropía

Un filtro de Kalman Básico difiere de un Filtro de Kalman de Máxima Correntropía (FKMCC) en su función de costo. Mientras que para derivar al primero se utiliza el error cuadrático medio (Ver Ecuación 2.5), el segundo utiliza el criterio de máxima correntropía. La derivación del FKMCC se muestra a continuación:

$$\begin{bmatrix} \hat{x}(k|k-1) \\ y(k) \end{bmatrix} = \begin{bmatrix} I \\ H(k) \end{bmatrix} x(k) + v(k), \quad (2.15)$$

con

$$v(k) = \begin{bmatrix} \hat{x}(k|k-1) - x(k) \\ r(k) \end{bmatrix}, \quad (2.16)$$

donde

$$E(v(k)v^T(k)) = \begin{bmatrix} P(k|k-1) & 0 \\ 0 & R(k) \end{bmatrix}. \quad (2.17)$$

Utilizando la descomposición de Cholesky, la ecuación 2.17 queda como sigue,

$$B(k)B^T(k) = \begin{bmatrix} B_p(k|k-1)B_p^T(k|k-1) & 0 \\ 0 & B_R(k)B_R^T(k) \end{bmatrix}, \quad (2.18)$$

donde $B(k)$ corresponde a dicha descomposición del término $E(v(k)v^T(k))$, $B_R(k)$ corresponde a la covarianza asociada al ruido de la medición $R(k)$ y $B_p(k)$ corresponde a la covarianza asociada al estado estimado.

Luego, al multiplicar por B^{-1} la ecuación 2.15 se obtiene,

$$D(k) = W(k)x(k) + e(k), \quad (2.19)$$

donde

$$D(k) = B^{-1} \begin{bmatrix} \hat{x}(k|k-1) \\ y(k) \end{bmatrix}, W(k) = B^{-1} \begin{bmatrix} I \\ H(k) \end{bmatrix}, e(k) = B^{-1}v(k).$$

Una vez realizado lo anterior, se propone una función de costo basada en el criterio de MCC,

$$J_L(x) = \frac{1}{L} \sum_{i=1}^L G_\sigma(d_i(k) - w_i(k)x(k)), \quad (2.20)$$

donde $d_i(k)$ es el i -ésimo elemento de $D(k)$, $w_i(k)$ corresponde a la i -ésima columna de $W(k)$ y L , que corresponde a la suma de las dimensiones de los vectores de estado de la observación (m) y del modelo (n), es la dimensión de $D(k)$.

Entonces, bajo el criterio de máxima correntropía es posible encontrar el estado óptimo $x(k)$ a partir de:

$$\hat{x} = \arg \max_x J_L(x) = \arg \max_x \sum_{i=1}^L G_\sigma(e_i(k)), \quad (2.21)$$

donde

$$e_i(k) = d_i(k) - w_i(k)x(k). \quad (2.22)$$

De forma análoga al filtro de Kalman básico, la solución óptima se determina a través de la derivada del costo con respecto al estado,

$$\frac{\partial J_L(k)}{\partial x(k)} = 0 \quad (2.23)$$

Luego, de esto resulta la siguiente ecuación:

$$x(k) = \left(\sum_{i=1}^L [G_\sigma(e_i(k))w_i^T(k)w_i(k)] \right)^{-1} \times \left(\sum_{i=1}^L [G_\sigma(e_i(k))w_i^T(k)d_i(k)] \right). \quad (2.24)$$

Debido a que $e_i(k) = d_i(k) - w_i(k)x(k)$ depende de $x(k)$, la solución óptima mostrada en la ecuación 2.24 corresponde a una ecuación de punto fijo de $x(k)$ la cual puede ser reescrita como sigue:

$$x(k) = f(x(k)), \quad (2.25)$$

donde

$$f(x(k)) = \left(\sum_{i=1}^L [G_\sigma(d_i(k) - w_i(k)x(k))w_i^T(k)w_i(k)] \right)^{-1} \times \left(\sum_{i=1}^L [G_\sigma(d_i(k) - w_i(k)x(k))w_i^T(k)d_i(k)] \right). \quad (2.26)$$

Un algoritmo de punto fijo iterativo puede ser obtenido a partir de:

$$\hat{x}(k)_{t+1} = f(\hat{x}(k))_t, \quad (2.27)$$

donde $\hat{x}(k)_t$ corresponde a la solución de la iteración del punto fijo en t.

La ecuación 2.24 puede ser reescrita como:

$$x(k) = (W^T(k)C(k)W(k))^{-1}W^T(k)C(k)D(k), \quad (2.28)$$

donde $C(k) = \begin{bmatrix} C_x(k) & 0 \\ 0 & C_y(k) \end{bmatrix}$ con

$$C_x(k) = \text{diag}(G_\sigma(e_1(k)), \dots, G_\sigma(e_n(k))) \text{ y } C_y(k) = \text{diag}(G_\sigma(e_{n+1}(k)), \dots, G_\sigma(e_{n+m}(k))).$$

La ecuación 2.28 puede ser expresada de la siguiente forma:

$$x(k) = \hat{x}(k|k-1) + \bar{K}(k)(y(k) - H(k)\hat{x}(k|k-1)), \quad (2.29)$$

donde

$$\bar{K}(k) = \bar{P}(k|k-1)H^T(k)(H(k)\bar{P}(k|k-1)H^T(k) + \bar{R}(k))^{-1}, \quad (2.30)$$

$$\bar{P}(k|k-1) = B_p(k|k-1)C_x^{-1}(k)B_p^T(k|k-1), \quad (2.31)$$

$$\bar{R}(k) = B_r(k)C_y^{-1}(k)B_r^T(k). \quad (2.32)$$

Se tiene que a mayor distancia la medición se trataría de un *outlier* por lo cual esta poseerá una baja influencia a la hora de estimar el estado, pues el valor se encontraría en la cola del *kernel* Gaussiano. Por otro lado, si la distancia es pequeña, entonces se puede confiar en la medida, pues el argumento del *kernel* Gaussiano se encontraría cercano a su *peak*.

El parámetro que establece si la distancia es grande o pequeña corresponde ancho de banda del *kernel* σ , el cual es esencial para que el Filtro de Máxima Correntropía funcione correctamente, es por esto que se analizará la sensibilidad de éste a través de métodos de ensayo y error como se propone en [10].

2.5. Programa

A continuación se describe el funcionamiento del programa elaborado por trabajos anteriores, explicando los elementos que recibe como entrada, qué es lo que hace con ellos y finalmente sus salidas. El detalle del código se encuentra en el trabajo de título de Paloma Pérez [7].

2.5.1. Entradas

En este apartado se mencionan y describen brevemente los elementos utilizados como entradas del programa original.

Los primeros elementos que recibe corresponden a la **imagen científica**, la cual es una imagen preprocesada, limpiada de rayos cósmicos y alineada con respecto a la imagen base, la **imagen de diferencia**, que corresponde a la diferencia entre la imagen científica y otra de referencia, la **imagen de varianza de la diferencia** y una *Point Spread Function empírica* (PSF ¹) que son derivadas por HiTS.

A través de la técnica de fotometría óptima de Naylor [12] se utilizan los elementos anteriores para conseguir 2 imágenes nuevas, las cuales corresponden a **imagen de flujo** e **imagen de varianza de flujo**. Estos nuevos elementos serán utilizados como entradas en los filtros de Kalman, los que corregirán la información predicha por un modelo matemático que describe las variables de estado: flujo y velocidad de flujo. Finalmente, el filtro arrojará 4 imágenes nuevas: **imagen de flujo estimado**, **imagen de velocidad de flujo estimado**, **imagen de varianza de flujo estimado**, **imagen de varianza de velocidad de flujo estimado**.

La última información que será considerada por el programa es la **imagen máscara** la cual señala los píxeles defectuosos que no deberían ser considerados para el análisis.

Las imágenes y elementos mencionados anteriormente a partir de ahora se denominan **serie de imágenes**, las cuales serán obtenidos por cada época de la que se tenga información. Un resumen de éstas se observan en el siguiente listado:

- Imagen científica
- Imagen de diferencia
- Imagen de varianza de la diferencia
- PSF empírica
- Imagen de flujo

¹Respuesta de un telescopio a una fuente puntual de luz. La luz de esta fuente es distorsionada por la atmósfera terrestre y los lentes del telescopio.

- Imagen de varianza de flujo
- Imagen de flujo estimado
- Imagen de velocidad de flujo estimado
- Imagen de varianza de flujo estimado
- Imagen de varianza de velocidad de flujo estimado
- Imagen de máscara.

2.5.2. Etapas de filtrado

El programa trabaja sobre un campo (*field*) y un CCD específico en el que se analizan iterativamente las series de imágenes por cada época de la que se tenga información.

Iterativamente, el programa revisa qué píxeles y grupo de píxeles por época cumplen con una serie de condiciones (en total 16) para ser considerado como parte o candidato a supernova. Las condiciones se separan en dos grupos:

El primer grupo es el denominado **filtro por píxeles individuales** que consta de 7 condiciones que deben satisfacer las imágenes de entrada correspondiente a la época en estudio para poder considerar el píxel como “píxel candidato”. Además, para que un píxel pueda ser considerado como candidato se requiere que éste sea detectado durante n épocas consecutivas, donde en este caso se utilizó el mismo valor que los autores anteriores, es decir $n = 4$. Una de estas condiciones analiza el valor del flujo estimado por el filtro de Kalman de máxima correntropía, donde se analiza qué píxeles superan un umbral impuesto. Si el umbral es superado, entonces el píxel podría ser considerado como candidato. Para ser considerado como un “píxel candidato” el mismo píxel debe superar las otras 6 condiciones impuestas.

Todos los píxeles que sean considerados como “píxel candidato” pasarán a una segunda etapa, **agrupamiento de píxeles**, en donde se agruparan píxeles según su cercanía. Estos grupos de píxeles serían considerados como “pre candidatos” a supernovas.

Finalmente, para transformar estos “pre candidatos” a candidatos, los grupos tendrán que satisfacer 9 condiciones más, las cuales son denominadas **filtros grupales**. Estas condiciones se basan en analizar qué es lo que ocurre alrededor de los centros de los grupos encontrados en el paso de agrupamiento de píxeles en las distintas imágenes mencionadas anteriormente. Una de estas condiciones se encarga de determinar si alrededor del píxel central en estudio existen píxeles defectuosos, en caso de que esto ocurra el grupo será descartado.

Una representación del algoritmo descrito anteriormente puede ser observado en la Figura 2.7.

2.5.3. Resultados que se obtienen

El programa emite una alerta de detección si la serie de imágenes correspondiente a una época supera todas las condiciones impuestas. Esta alerta entregará la posición del píxel central de la supernova junto a la época en la que fue detectada la supernova.

Además de entregar la posición y la época en la que fue detectada, se obtienen imágenes,

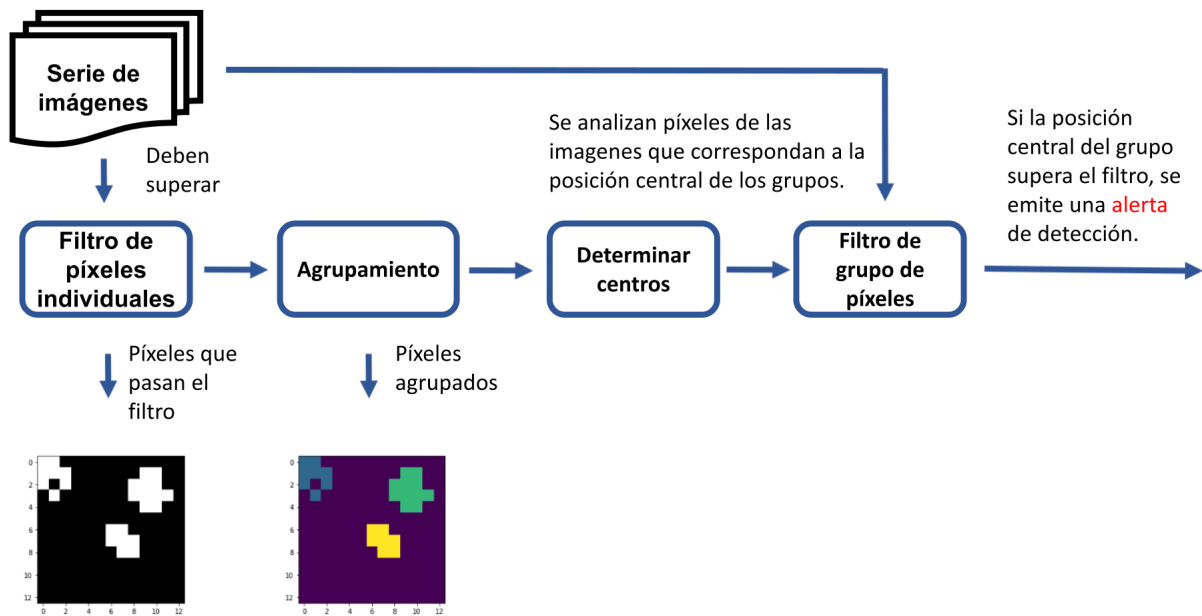


Figura 2.7: Representación de funcionamiento de la rutina propuesta por autores anteriores para detectar supernovas. El programa inicia leyendo la serie de imágenes correspondientes a una época en particular, donde se analiza si los píxeles individuales de cada una de estas cumple con el **Filtro de píxeles individuales**. Los píxeles que superan las 7 condiciones durante 4 épocas consecutivas pasarán a una etapa de **Agrupamiento**, donde se agruparan los píxeles según su cercanía. A partir de los nuevos grupos, se determinarán sus centros para ser filtrados por el **Filtro de grupo de píxeles**. Finalmente, si la serie de imágenes y los píxeles centrales cumplen con 9 nuevas reglas se emitirá una alerta.

que serán mostradas en el capítulo 4 de Resultados y Análisis en la sección 4.2, las cuales son utilizadas para analizar el comportamiento de las modificaciones que se realicen.

2.6. Laboratorio Nacional de Computación de Alto Desempeño (NLHPC)

El *National Laboratory for High Performance Computing* (NLHPC) es un proyecto financiado por el PIA de CONICYT, alojado en el Centro de Modelamiento Matemático (CMM) en la escuela de Ingeniería y Ciencias (FCFM), Universidad de Chile, orientado a la instalación y mantención de una capacidad computacional capaz de satisfacer las demandas de computación de alto rendimiento nacional. Éste ofrece servicios de alta calidad y promueve su uso tanto en investigaciones básicas como en aplicaciones industriales.

Las características de Leftraru, uno de sus *clusters* que posee la base de datos obtenidas por HiTS, se muestran a continuación:

- 132 Nodos de cómputo HP (128 nodos HP SL230 y 4 nodos HP SL250), cada uno con 2 procesadores de 10 cores Intel Xeon Ivy Bridge E5-2660 V2.
- 2640 cores.

- 12 Co-procesadores Intel Xeon Phi 5110p de 2 TFlops.
- 6.25 TB de RAM
- Conexión Infiniband FDR a 56Gbps
- 274TB de almacenamiento Lustre (DDN EXAScaler)
- Capacidad de cómputo: 70 TFlops.

2.7. Generalización de la entropía para el análisis de curvas

En [11] se propone un método con el cual es posible analizar y eventualmente clasificar sistemas dinámicos. Este método además resulta ser una extensión de la entropía de las curvas de un plano al espacio R^n .

Las propiedades de un sistema dinámico se pueden inferir a partir del análisis de la evolución de la curva de estados de un sistema. Si la curva en estudio es una línea recta, entonces la entropía de ésta será 0. Por otro lado, si la curva posee un movimiento más desordenado u ondulado su entropía será mayor.

El cálculo de la entropía de una curva plana se realiza a partir de la siguiente formulación:

$$H(\Gamma) = \log \left(\frac{2L}{C} \right) \quad (2.33)$$

donde Γ es la representación de la curva, L al largo de la curva Γ y C es la longitud del límite de su casco convexo (*convex hull*), la cual es la envolvente convexa de un conjunto de puntos de una dimensión n .

Para este trabajo se utilizarán las curvas en espacio de estado de las supernovas, obtenidas a partir de las estimaciones de los estados que arroje el filtro de Kalman, donde se espera que los objetos que no son supernovas (*outliers* u otro objeto) posean una mayor entropía que los que si lo sean.

Capítulo 3

Metodología

En el presente capítulo se presenta la metodología usada para cumplir con los objetivos del trabajo. Este capítulo se inicia con una descripción generalizada, donde se explica la estructura de los experimentos a realizar, los que luego son explicados y detallados en subsecciones.

3.1. Metodología general

Para empezar, se lleva a cabo un relajamiento de los parámetros asociados a las condiciones del programa original, las cuales se encuentran descritas en Apéndice A.1 y Apéndice A.2 y en la sección de Apéndice de la memoria de Paloma Pérez [7], a pesar de esto, en el desarrollo de este trabajo se describirán las que se utilizarán. Lo anterior es realizado para determinar si es posible encontrar más supernovas confirmadas por HiTS, manteniendo un número manejable de falsos positivos. Los parámetros serán modificados si es que estos no aumentan la cantidad de falsos a números inmanejables y si son capaces de aumentar la cantidad inicial de supernovas detectadas.

Luego, se propone utilizar el cálculo de entropía de curvas y el ajuste paramétrico del filtro de Kalman de máxima correntropía para discriminar los candidatos obtenidos en la parte anterior, donde se espera descartar la mayor cantidad de falsos positivos manteniendo un alto número de detecciones correctas.

A continuación, se realiza una poda de condiciones que no aporten en la discriminación de supernovas del modelo resultante del paso anterior. En caso de ser necesario se volverán a apretar y/o relajar condiciones.

Cada una de las etapas anteriores serán comparadas con el modelo original, considerando:

- Número de verdaderos positivos y falsos positivos.
- Cantidad de veces que los modelos detectan las mismas supernovas.
- Qué tan temprano se detectan las mismas supernovas, comparándolas con la campaña de HiTS y los resultados anteriores.

Lo descrito anteriormente es resumido en el diagrama de la Figura 3.1, donde se muestra la acción y el objetivo de cada punto.

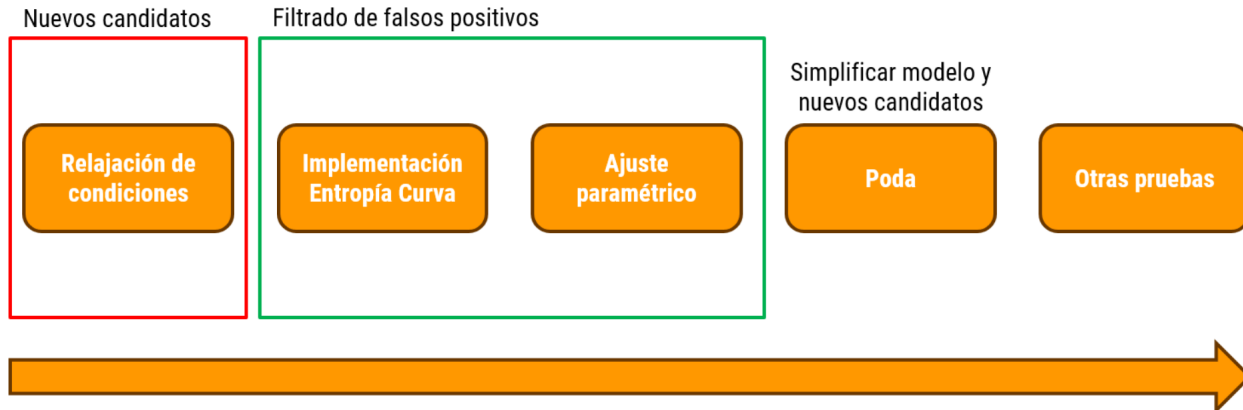


Figura 3.1: Diagrama que representa la metodología utilizada en este trabajo. Se inicializa relajando las condiciones del programa original para obtener nuevos candidatos, luego se implementa una nueva condición y se ajustan parámetros para filtrar los falsos positivos obtenidos del paso anterior, finalmente se procede a podar y a realizar otras pruebas que serán mencionadas más adelante.

3.2. Base de datos

Se trabaja con las series de imágenes provenientes de las campañas del sondeo HiTS de los años 2014 y 2015. La campaña del 2015A reportó 90 candidatos a supernovas en 89 series de imágenes distintas (se detectaron 2 supernovas dentro de la misma serie). Estas 89 series de imágenes serán repartidas en 3 conjuntos, entrenamiento, validación y prueba.

- Set de entrenamiento: Utilizado para entrenar el modelo.
- Set de validación: Utilizado para ajustar los parámetros de modelo.
- Set de prueba: Utilizado para determinar qué tan bien generaliza el modelo propuesto. Estos datos son completamente nuevos y no han interferido de ningún modo en el ajuste y/o elección de los parámetros y/o modelos.

La campaña del 2014A reportó 34 candidatos a supernovas en 34 series de imágenes distintas. Esta serie se utilizará como conjunto de prueba adicional para determinar qué tan bien generalizan los modelos finales ante series de imágenes que presentan características distintas. Estas series fueron expuestas por un largo periodo de tiempo y poseen cadencias más lentas en comparación a las series de la campaña del 2015.

La cadencia de observaciones de las dos campañas mencionadas anteriormente se pueden apreciar en las Figuras 3.2 y 3.3.

3.3. Modelo de trabajos anteriores

Para iniciar el análisis, se utiliza el modelo propuesto por los autores anteriores en la base de datos de HiTS 2015A, cuyos resultados son comparados con las nuevas implementaciones

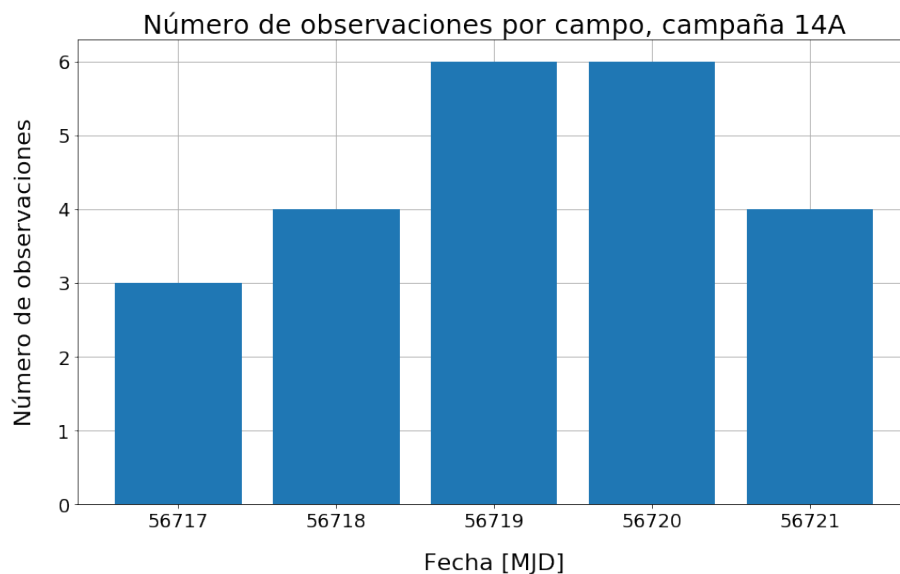


Figura 3.2: Cadencia de observaciones de campaña 2014A (5 noches).

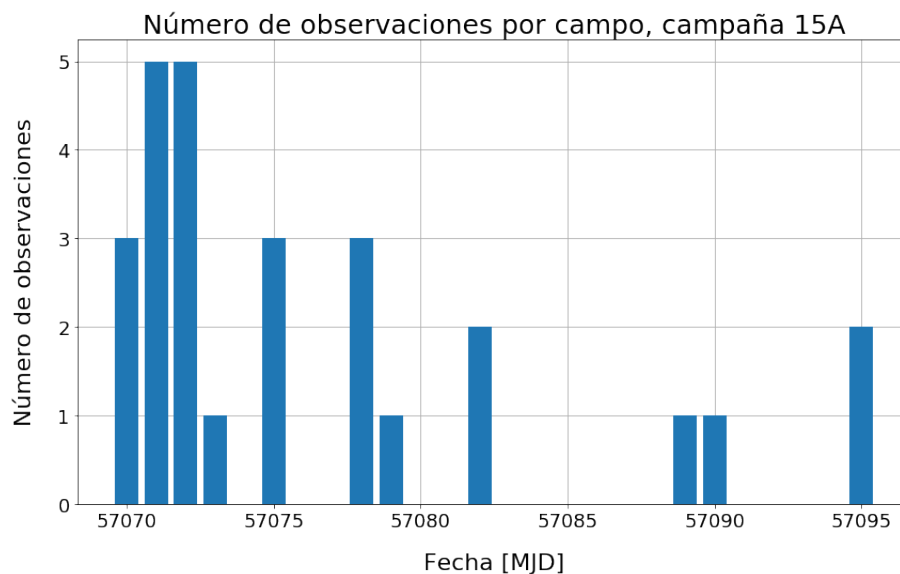


Figura 3.3: Cadencia de observaciones de campaña 2015A (11 noches).

y ajustes establecidos. Este modelo se denominará **modelo original**.

3.4. *Zero points*

Un *zero point* es un factor que posee información de la absorción y correcciones de la difracción del color producidas por la atmósfera. Este valor es calculado para hacer coincidir la magnitud aparente de una estrella, es decir, el brillo medido desde la Tierra de la fuente, con una magnitud de referencia dada. A partir de lo anterior, se implementará este factor con el objetivo de corregir las curvas de flujo que se obtienen a través de la fotometría de Naylor. En particular, se implementa el factor γ obtenido por Rodrigo Carrasco en su trabajo de tesis [14], el cual permite, a través de la multiplicación de dicho factor con la imagen científica

preprocesada, obtener la información del flujo en ADU ¹ corregida.

$$\gamma = 10^{\frac{Z_p^{ref} - Z_p(t)}{2.5}}, \quad (3.1)$$

donde Z_p^{ref} es el *zero point* utilizado como referencia y $Z_p(t)$ es el *zero point* correspondiente a la época en estudio.

Se compara el modelo original con y sin este factor de corrección.

3.5. Relajación de condiciones

El objetivo de esta sección es encontrar una mayor cantidad de supernovas (redescubiertas por HiTS) a través del relajamiento de parámetros asociados a las 16 condiciones del programa original. Se espera que al realizar dicho relajamiento, aumente la cantidad de falsos positivos, que serán abordados y/o discriminados a través de una condición asociada al cálculo de entropía de curvas y al ajuste paramétrico del filtro de Kalman de máxima correntropía.

Además se estudia el efecto que tienen estos parámetros sobre el programa y las relaciones que existe entre las condiciones de la etapa de filtrado de píxeles individuales y las de la etapa de filtrado por grupos. Se utilizarán los valores de los parámetros que aumenten la cantidad de detecciones correctas y no provoquen aumentos inmanejables de falsos positivos.

3.5.1. Variables acumuladoras

Para empezar, se exploran dos parámetros denominados: **accum neg flux depth** y **accum med flux depth**.

El primero (**accum neg flux depth**) se relaciona con una variable denominada **accum neg flux**, la cual determina qué píxeles poseen valores negativos de flujo. Si el píxel no presenta un valor negativo durante “**accum neg flux depth**” imágenes consecutivas entonces éste es considerado como un píxel candidato, en caso contrario, este píxel será descartado. Este parámetro está asociado a una condición de filtro por grupos, la cual se encarga de descartar grupos de píxeles candidatos que posean valores negativos al su alrededor.

El segundo (**accum med flux depth**) se relaciona con una variable denominada **accum med flux**, la cual está enfocada en guardar “**accum med flux depth**” valores de flujo de cada píxel para luego calcular su mediana. Los píxeles que posean una mediana de flujo superior a un cierto umbral (en este caso 1500) serían descartados. Este parámetro está asociado a una condición de filtro de píxeles individuales, la cual descarta píxeles individuales que no cumplan con la condición anterior, siendo considerados como píxeles defectuosos.

Se realizan pruebas utilizando los siguientes valores:

- **Accum neg flux depth** = 4 y **accum med flux depth** = 3. (Modelo original)

¹ *analog to digital unit*

- Accum neg flux depth = 1 y accum med flux depth = 1.

Se busca determinar si utilizando el caso extremo (llevando estos parámetros a 1), se logra determinar una mayor cantidad de supernovas redescubiertas manteniendo un número manejable de falsos positivos, y en caso de que esto no ocurra, se estudiarán más combinaciones de estos parámetros.

3.5.2. Umbrales de flujo y velocidad de flujo

Los parámetros a explorar son: **Umbral de flujo** y **Umbral de velocidad de flujo**.

Estos parámetros pertenecen a la etapa de filtrado de píxeles individuales, donde se comparan los resultados de los estados flujo y velocidad de flujo estimados, es decir, los obtenidos por el filtro de Kalman de máxima correntropía. Cada píxel individualmente debe superar estos umbrales para ser considerado como un píxel válido o píxel candidato.

Estos parámetros son estudiados pues la relajación de estos podría generar detecciones más tempranas de supernovas, dado que si estas son observadas en sus primeros momentos, su flujo y velocidad de flujo podrían pasar desapercibidos por otros métodos, que para evitar confundirse con ruido, establecen umbrales de flujo altos. Por esto mismo, los parámetros no pueden ser demasiado bajos dado que se podría confundir el comportamiento temprano de una supernova con el ruido de observación.

Los valores con los que se estudiarán dichos parámetros son:

- Flujo: 200 [ADU] (modelo original), 175 [ADU], 150 [ADU].
- Velocidad de flujo: 50 [ADU/días] (modelo original), 40 [ADU/días], 30 [ADU/días], 20 [ADU/días], 10 [ADU/días], 1 [ADU/días].

3.5.3. ASMR

El parámetro a explorar es el denominado **ASMR**, el cual significa *Above local Science Median*.

Este parámetro está asociado a una condición de la etapa de filtrado por grupos. Dicha condición se encarga de verificar si los píxeles dentro de un área, que rodea al píxel central del grupo de píxeles, superan la mediana local en la imagen científica.

Este parámetro ajusta el área a considerar para calcular la mediana local y compararla con la de la imagen científica, donde mientras más alto sea este valor se tomará un mayor área alrededor del píxel de interés. Se busca explorar si es que al tomar un área aún mayor es posible aumentar la cantidad de detecciones.

Este parámetro será variado en:

- ASMR: 3 (modelo original), 10, 20, 30.

3.5.4. Varianza de flujo

El parámetro a explorar es: **varianza de flujo**.

Este parámetro está asociado a una condición de la última etapa de filtrado, filtrado por grupo, la cual está encargada de determinar si el valor de la varianza de flujo del píxel central de un grupo de píxeles es superior al umbral “varianza de flujo”.

Este parámetro será variado en:

- Varianza de flujo: 250 [ADU^2] (modelo original) y 500 [ADU^2].

3.5.5. Modelo relajado

El modelo obtenido a partir del relajamiento de los parámetros antes mencionados, se denominará **modelo relajado**, el cual será comparado con el modelo original a través del análisis de: las detecciones correctas de supernovas, qué tan temprano, con respecto a HiTS, son detectadas y cuántas veces son detectadas por estos dos modelos.

3.6. Entropía de curvas

El objetivo de esta sección es descartar, a través de la implementación del cálculo de la entropía de curvas, la mayor cantidad de falsos positivos manteniendo la mayor cantidad de verdaderos positivos.

La implementación de esta nueva condición no interviene la rutina del modelo de trabajos anteriores, sino más bien, filtra las alertas obtenidas por éste.

La curva a estudiar es la compuesta por los valores de flujo estimado y velocidad de flujo estimado del píxel central del grupo de interés, es decir, a ésta se le calcula la entropía. En la Figura 3.4 se observa un ejemplo de ésta.

Según la Ecuación 3.1, para calcular la entropía de la curva se requiere conocer su largo y el perímetro de su *convex hull*. Para obtener este último elemento se siguen los siguientes pasos:

- Primero: Se determinan los puntos de la curva. Ver Figura 3.5a.
- Segundo: Se identifican los vértices de la envolvente convexa. Ver Figura 3.5b.
- Tercero: Se unen estos vértices para formar la envolvente convexa. Ver Figura 3.5c.

El método consiste en calcular la entropía de la curva en el espacio de estados considerando el punto donde el modelo original emitía una alerta de supernova y los N puntos anteriores de las N épocas anteriores. Si la entropía de la curva supera un cierto umbral, la alerta es descartada, pero si por el contrario está bajo el umbral, la alerta es considerada como verdadera.

Es importante señalar que el cálculo de la entropía de las curvas considera al menos 4 puntos y como máximo una cantidad “lim” de puntos.

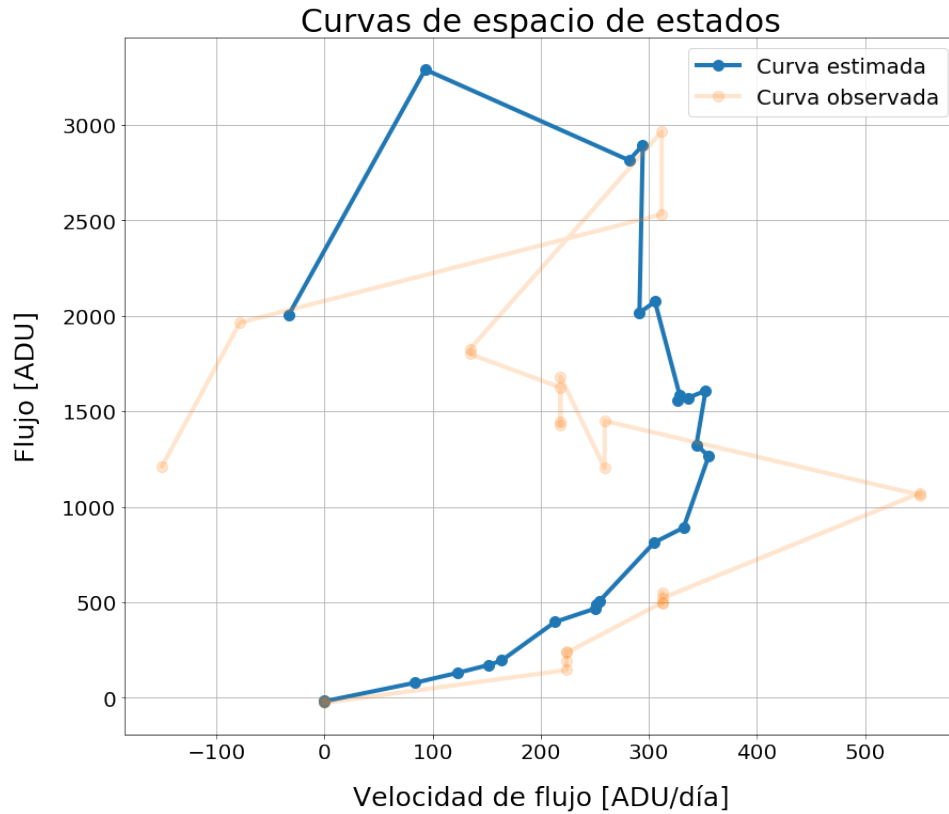


Figura 3.4: Curva de espacio de estados de la supernova encontrada en el campo 02 y CCD N5. En azul se observa al curva estimada por el filtro de Kalman de máxima correntropía y en rojo se observa la curva de los valores observados.

Los valores con los que se estudiarán ambos parámetros: (umbral de entropía y “lim”) son:

- umbral de entropía: sin condición, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9 y 1.0
- lim: 6, 10, 14, 16 y 18 puntos

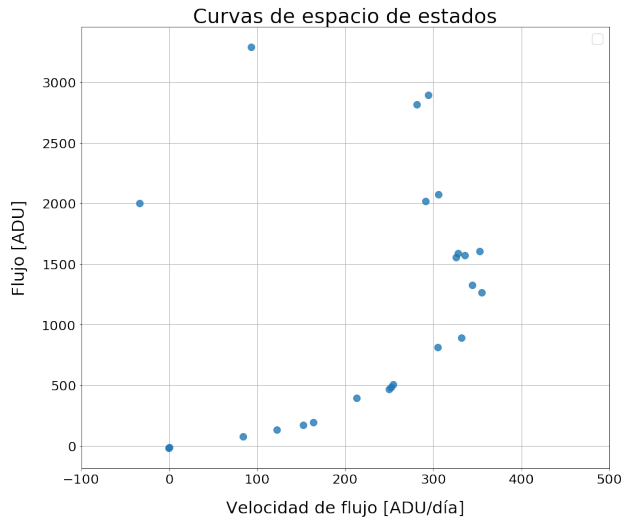
Con respecto al segundo parámetro, se escogerá el menor valor posible donde se pueda discriminar la mayor cantidad de falsos positivos sin perder verdaderos positivos.

3.7. Filtro

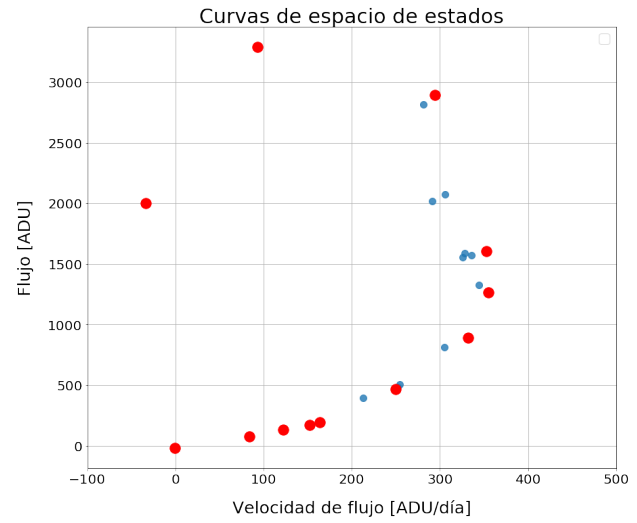
En esta sección se pretende estudiar distintos metaparámetros relacionados al filtro de Kalman de máxima correntropía, con el fin de ajustar aún más el modelo obtenido del punto anterior.

Los parámetros que serán explorados son:

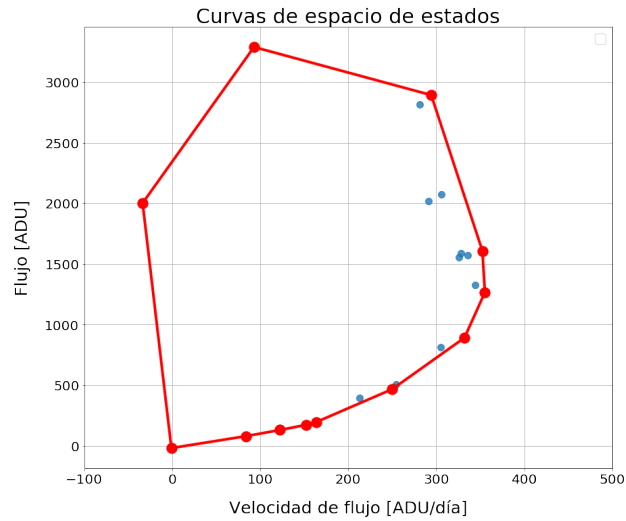
- σ_{modelo} : Este parámetro corresponde a la desviación estándar del ruido del proceso, el cual es un factor con el que es posible manejar el grado de incerteza que tiene el modelo según la diferencia de tiempo que existe entre observaciones. Se espera que mientras mayor sea la diferencia de tiempo entre una observación y otra, el modelo matemático



(a) Puntos curva de espacio de estados



(b) Vértices de polígono



(c) Convex Hull

Figura 3.5: Etapas para obtener el *convex hull*. (a) Muestra los puntos de la curva en el espacio de estados de una supernova, (b) muestra los vértices de la envolvente convexa y (c) muestra la envolvente convexa.

fallará en predecir correctamente el siguiente punto, por lo que se le debe dar una mayor confianza a la observación en vez del modelo.

- $\sigma_{\text{correntropia}}$: Parámetro del filtro de máxima correntropía y corresponde al ancho de banda del *kernel* Gaussiano utilizado para la estimación de la correntropía. Se espera que un ajuste adecuado sea capaz de filtrar *outliers* presentes en las observaciones.

El mejor modelo obtenido en esta sección se denominará **modelo final**.

3.8. Poda

En esta etapa se realiza una poda del programa, es decir, se eliminan condiciones que no muestren un aporte significativo en los resultados. Se espera obtener como principales ventajas:

- Simplificar el programa, lo que facilitará su estudio paramétrico.
- Evitar sobreajuste.
- Aumentar la cantidad de detecciones correctas.
- Lograr detectar numerosas veces una misma supernova.
- Detectar tempranamente las supernovas.

Para realizar esta poda se eliminan progresivamente distintas condiciones. Si al eliminar una condición se observa que no existen grandes cambios, entonces se procede a mantener dicha condición eliminada y se estudia el comportamiento de eliminar una segunda condición. Esto se realiza hasta que al eliminar alguna condición más, el número de falsos positivos aumente considerablemente. El modelo resultante de este experimento se denominará *modelo final podado*.

3.9. Otras pruebas

3.9.1. Comparación de modelos obtenidos en distintas secciones

En esta sub sección se comparan 4 de los modelos obtenidos a lo largo de los experimentos realizados. Dentro de las comparaciones se incluyen la cantidad de detecciones realizadas, la cantidad de veces y qué tan temprano son descubiertas las supernovas. Se mostrarán ejemplos de cómo alteran los distintos modelos a la etapa de definir los píxeles candidatos.

3.9.2. Alta cadencia *versus* baja cadencia

El filtro de Kalman puede provocar malas estimaciones si la diferencia de tiempo entre un instante y otro es muy grande, dado que se pierde información del comportamiento del sistema. Para comprobar si el modelo es capaz de lidiar contra este problema se propone estudiar el modelo final obtenido solo en su periodo de mayor cadencia y compararlo con los resultados del mismo pero considerando todas las observaciones.

Se espera que los métodos explorados en este trabajo luego sean utilizados en otras bases de datos, como en los futuros telescopios mapearán el cielo con alta cadencia. Es por esta

razón, que tiene sentido explorar el comportamiento del programa solo en periodos de alta cadencia.

3.9.3. Imágenes filtradas según *airmass*

La rutina original filtra las series de imágenes respectivas a una época cuya imagen científica superen el umbral *airmass*² impuesto, el cual en este caso es 1,7. El objetivo de este filtro es que si la imagen supera ese valor la observación no sería representativa y solo sería considerada como un *outlier*.

A través del modelo final encontrado se procederá a estudiar qué es lo que ocurre si se consideran todas las imágenes (incluyendo las que superen el umbral), donde se espera que el filtro de Kalman de máxima correntropía sea capaz de filtrar estos puntos, manteniendo así resultados similares a los que se obtienen considerando el umbral de *airmass*.

3.9.4. Comparación entre el uso de filtros de Kalman de máxima correntropía y básico

En este experimento se comparan los resultados del modelo final al variar su filtro por el filtro de Kalman básico y el de máxima correntropía. De este modo se podrá determinar ventajas y desventajas de usar el criterio de máxima correntropía.

3.9.5. Prueba con otra base de datos

Se utiliza el modelo original, el final y el final podado sobre un nuevo conjunto de datos provenientes de la campaña de HiTS del año 2014A. A pesar de que esta base de datos posee una menor cantidad de puntos, su exposición fue más larga que la de la campaña 2015A y su cadencia fue más lenta que ésta, se espera que los modelos sean capaces de encontrar una mayor cantidad de supernovas que el modelo original, manteniendo una cantidad de falsos positivos baja.

²Corresponde a la cantidad de atmósfera existente entre el telescopio y la estrella

Capítulo 4

Resultados y análisis

En este capítulo se muestran y analizan los resultados de los distintos experimentos explicados en la metodología. En cada sección se presentan tablas que muestran la cantidad de verdaderos positivos o TP (supernovas encontradas y confirmadas por HiTS), falsos positivos o FN (supernovas no encontradas por el programa pero que sí están confirmadas por HiTS) y falsos positivos o FP (elementos que no fueron clasificados como supernovas por HiTS) tanto del conjunto de entrenamiento como el de validación y, para el análisis de algunos modelos, el conjunto de prueba. Cabe señalar que dentro del grupo de falsos negativos pueden existir supernovas que HiTS no haya logrado detectar, es por esto que una vez obtenido un modelo que se considere mejor que el de trabajos anteriores, se analizarán los candidatos que el modelo final arroje.

Al final de cada sección se presenta un análisis de las ventajas y desventajas de cada modelo, donde además de considerar la cantidad de supernovas redescubiertas, se determinará qué tan temprano (con respecto a HiTS) y qué tantas veces son detectadas las supernovas confirmadas por HiTS, utilizando para esto el conjunto de prueba de la campaña de HiTS 2015A.

En la última sección se presentan los resultados de distintos experimentos, en donde se compararán los mejores modelos obtenidos de la sección. En esta parte se utilizan las series de imágenes provenientes de la campaña de HiTS 2014A y 2015A.

4.1. Base de datos

La base de datos compuesta por 89 series de imágenes obtenidas del sondeo HiTS de la campaña del 2015A será dividida de forma aleatoria en:

- 44 series para conjunto de entrenamiento (49,44 %).
- 22 series para conjunto de validación (24,72 %).
- 23 series para conjunto de pruebas (25,84 %).

Por otro lado, la base de datos compuesta por 32 series de imágenes obtenidas del sondeo

HiTS de la campaña del 2014A será utilizada solo como conjunto de prueba.

En cada serie de imágenes existe una supernova descubierta por HiTS.

4.2. Resultados de trabajos anteriores

Los resultados de ejecutar los modelos de trabajos anteriores se encuentran en la Tabla 4.1, donde es posible observar que no existen diferencias significativas en el conjunto de prueba (*test*) entre el filtro de Kalman básico y el de máxima correntropía (incluso no presentan diferencias significativas en los otros conjuntos), más aún, se detectan las mismas supernovas. Teóricamente esto puede ocurrir cuando el ancho de banda del *kernel* Gaussiano ($\sigma_{correntropia}$) es muy grande (tiende a infinito), lo que provoca que el filtro de Kalman de máxima correntropía se reduzca a un filtro de Kalman básico [10]. En la rutina original se utilizó un valor de $\sigma_{correntropia}$ muy alto ($\sigma_{correntropia} = 1000$), el cual provoca que los resultados entre los 2 filtros tiendan a parecerse.

De aquí en adelante, al modelo que utiliza el filtro de Kalman de máxima correntropía de los autores anteriores se denominará **modelo original**, el cual será utilizado para comparar el funcionamiento de los nuevos modelos propuestos en este trabajo. Principalmente, se trabajará con el filtro de Kalman de máxima correntropía dado que éste es capaz de manejar de mejor forma el ruido impulsivo [10], presente en los datos astronómicos, que el filtro de Kalman básico, dado que al ser derivado del error cuadrático medio es sensible a *outliers*. A pesar de esto, una vez obtenido un mejor modelo que el original, se compararán los resultados con el filtro de Kalman básico.

Tabla 4.1: Número de verdaderos positivos (TP), falsos negativos (FN) y falsos positivos (FP) detectados usando los filtros de Kalman básico y de máxima correntropía.

Filtro	Entrenamiento			Validación			Prueba		
	TP	FN	FP	TP	FN	FP	TP	FN	FP
Kalman básico	19	25	11	9	13	18	8	15	8
Max. correntropía	19	25	11	9	13	19	8	15	10

En la Figura 4.1 se aprecia una serie de gráficos que muestran información del píxel central de un candidato de supernova encontrado por el programa. En la Figura 4.2 se aprecia un gráfico que muestra la curva en el espacio de estados del píxel en estudio. En la Figura 4.3 se aprecia una serie de estampillas correspondientes a las cercanías del píxel en estudio. Las 3 figuras mencionadas anteriormente corresponden a versiones corregidas entregadas por la rutina original. Las versiones anteriores de estos gráficos pueden ser observadas en las Figuras 4.3, 4.4 y 4.5 respectivamente en la tesis de Paloma Pérez [7].

4.3. Zero points

Al incluir el factor de corrección γ (Ecuación 3.1) en el modelo original, no se observan diferencias significativas en la cantidad de detecciones correctas, pero sí en la cantidad de falsos positivos los cuales aumentan. (Lo anterior se observa en el Tabla B.1 presente en el

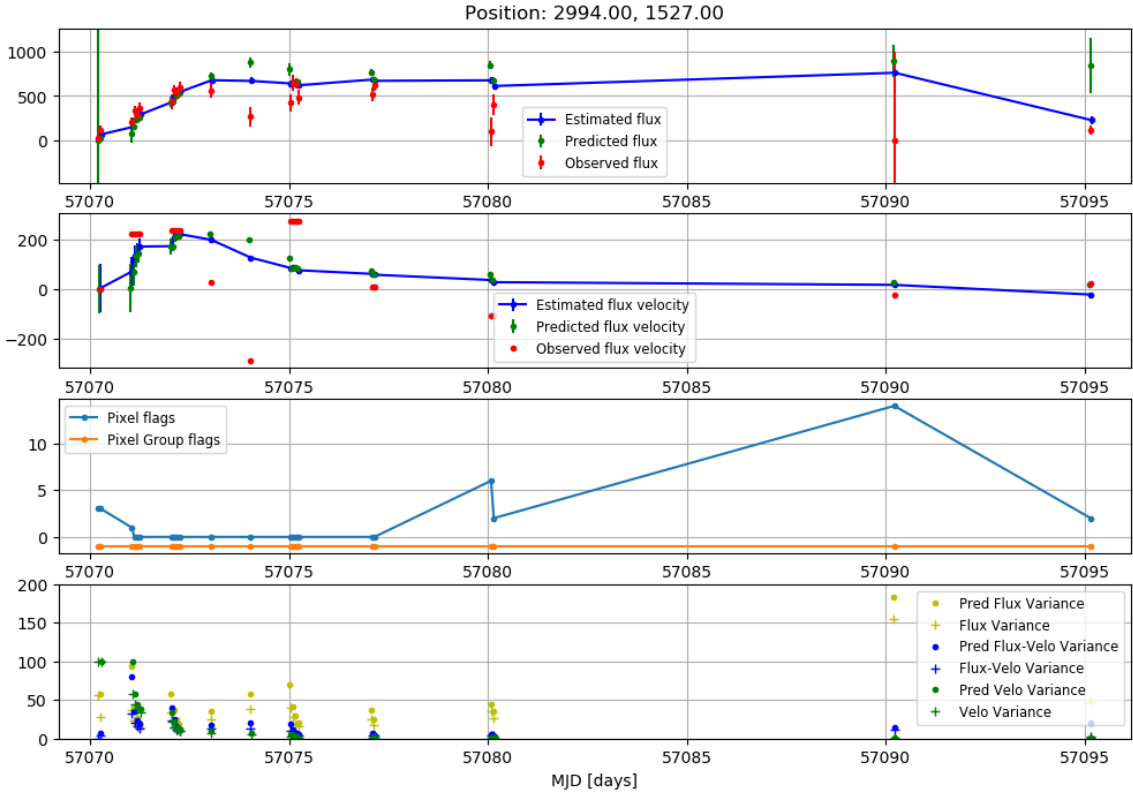


Figura 4.1: Conjunto de gráficos que muestran información referente al píxel en el que se emitió una alerta. De arriba hacia abajo se observa que el primer gráfico muestra el comportamiento del flujo en ADU a través del tiempo expresado en MJD, se muestra el estado observado (rojo), el estado predicho (verde) y estado estimado (azul) con sus respectivos errores. El segundo gráfico muestra la variable de estado de velocidad de flujo en ADU/días considerando su curva predicha (verde) y estimada (azul). El tercer gráfico muestra qué condiciones no fueron satisfechas por el píxel en estudio. Finalmente, en el último gráfico se muestra información sobre la varianza de las componentes predichas y estimadas.

apéndice B.1). Dado que éste es utilizado para corregir las curvas de flujo, se espera que los modelos futuros sean capaces de redescubrir una mayor cantidad de supernovas.

Dos ejemplos de curvas corregidas por el factor en estudio se muestran en las figuras Figura 4.4 y Figura 4.5 donde se aprecian las variaciones que presentan sus valores de flujo. Para la primera de las figuras mencionadas, el factor aumenta el valor de su flujo, lo cual permitiría que las observaciones pudieran superar más fácilmente los umbrales y por lo tanto lograr una detección más temprana. En la segunda imagen se muestra la curva de una supernova que anteriormente (sin el factor de corrección) no era detectada. En esta curva en particular se observa que la observación siguiente al *peak* presenta un claro aumento en el valor del flujo, dado que sin este factor, el flujo alcanzado era inferior a 500 [ADU] mientras que al considerarlo, éste aumentó a más de 1500 [ADU]. Lo anterior puede ocurrir cuando existen factores ambientales que reduzcan la cantidad de luz recibida por el telescopio.

A partir de ahora, este factor será incluido en todos los modelos, pues permite trabajar con una información más fidedigna de las curvas de luz.

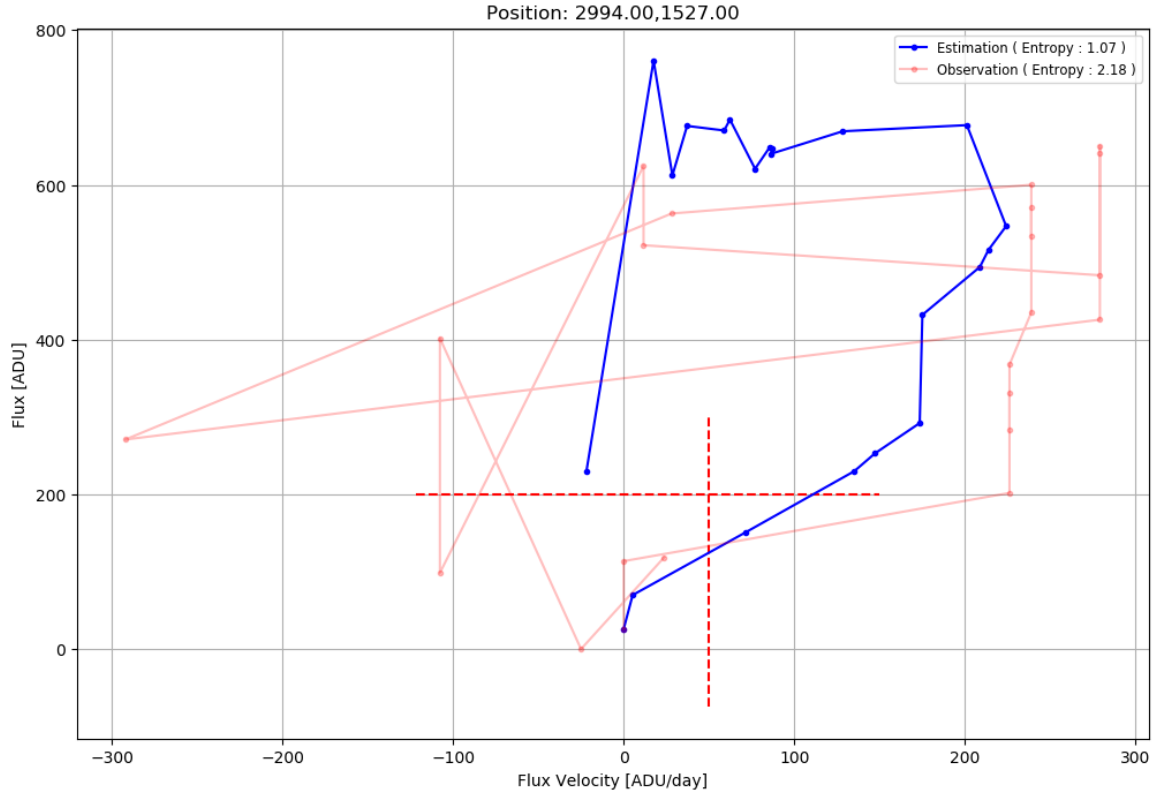


Figura 4.2: Gráfico que muestra la curva en el espacio estados del píxel en estudio. La curva estimada está representada en azul mientras que la observada en rojo. Se aprecia además unas líneas rectas discontinuas que corresponden a los umbrales que deben ser superados para que estos puntos sean considerados como candidatos. Cabe señalar que en la leyenda de arriba a la derecha se agrega la información de la entropía.

4.4. Relajación de condiciones

A continuación se presentan los resultados obtenidos al relajar los parámetros asociados a las condiciones del modelo de trabajos anteriores [6] y [7].

4.4.1. Variables acumuladoras

Estos parámetros permiten descartar píxeles y/o grupos de píxeles que presenten valores negativos en sus flujos. En este experimento se redujeron ambos valores a 1 con lo que se encontraron más detecciones correctas (TP) a un costo de obtener una mayor cantidad de falsos positivos (FP). La cantidad de candidatos de supernovas encontrados se pueden observar en la Tabla B.2 del Apéndice B.1.

La ventaja de disminuir estos valores a 1 es que la cantidad de píxeles individuales descartados disminuirá, dado que solo dependerá si el píxel o grupo de píxeles en estudio (de una época en particular) poseen o no, valores de flujo negativos y no considerará el comportamiento de dichos píxeles de épocas anteriores como lo hace el modelo inicial.

Debido a que estos valores son los más bajos que pueden ser utilizados para que las condi-

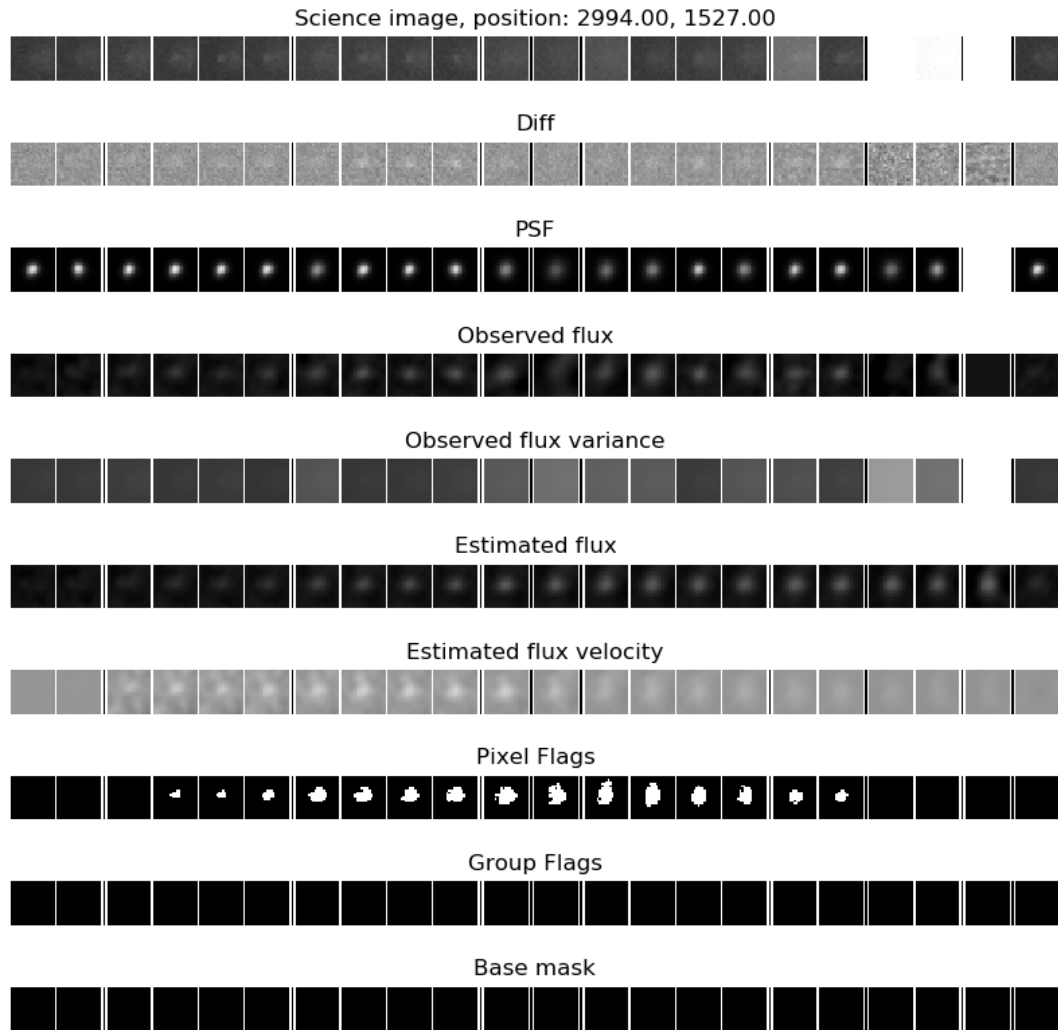


Figura 4.3: Serie de estampillas que muestran la información alrededor del píxel en estudio por época, donde se muestran (de arriba hacia abajo) las imágenes científicas, imágenes de diferencia, PSF, flujo observado, varianza del flujo observado, flujo estimado, velocidad de flujo estimada, grupo de píxeles candidatos alrededor del central, grupos que son descartados, píxeles que son defectuosos (*base mask*).

ciones no pierdan su sentido y además permiten detectar una mayor cantidad de supernovas sin que la cantidad de falsos positivos aumenten a números inmanejables, no se explorarán más combinaciones.

De aquí en adelante se trabajará con los parámetros iguales a 1.

4.4.2. Velocidad de flujo estimada

Un umbral relevante es el relacionado a la velocidad de flujo obtenida del filtro de Kalman de máxima correntropía. Se probó variando este umbral con valores positivos menores a 50, dado que se desea relajar el valor propuesto por trabajos anteriores. Se probaron con los valores 1, 10, 20, 30, 40 y 50, cuyos resultados se pueden ver en la Tabla 4.2. A partir de

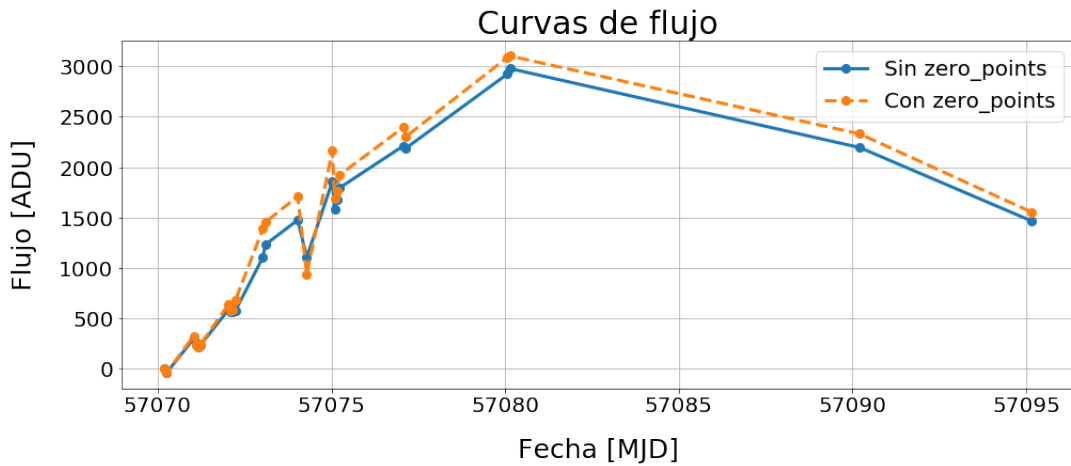


Figura 4.4: Comparación de curvas de flujo en ADU con y sin el factor de corrección de la supernova encontrada en el campo 02 y CCD N5.

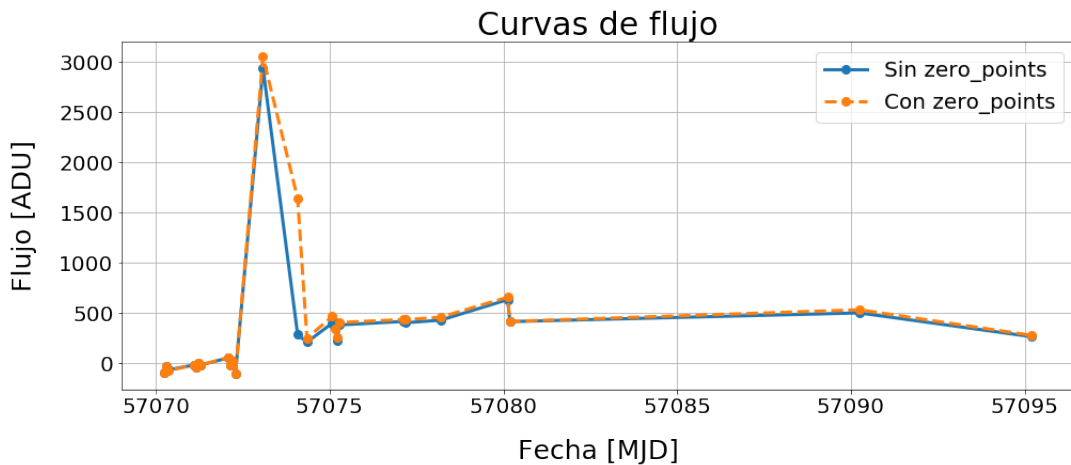


Figura 4.5: Comparación de curvas de flujo en ADU con y sin el factor de corrección de la supernova encontrada en el campo 28 y CCD S16.

la variación de este, se observa que el umbral con el que se obtiene la mayor cantidad de supernovas es el de 20.

Dado que se están relajando condiciones, se esperaría obtener una mayor cantidad de candidatos, manteniendo o aumentando la cantidad de detecciones correctas, pero esto no ocurre. En dicha tabla se observan variaciones en la cantidad de TP, donde al utilizar un umbral de 20 se obtiene la mayor cantidad de supernovas, pero luego, al seguir reduciéndolo, se pierden detecciones correctas. Lo anterior se debe a la estructura misma de la rutina original donde, como ya fue explicado anteriormente, la primera etapa es la de filtrado de píxeles individuales que luego son agrupados para finalmente ser filtrados de forma grupal. Dado que el umbral de velocidad de flujo pertenece al primer filtro, su relajación provocará que se descarten menos píxeles de la imagen provocando que la etapa de agrupamiento reúna elementos ruidosos que alterarán los resultados.

De aquí en adelante se trabajará con velocidad de flujo igual a 20.

Tabla 4.2: Número de verdaderos positivos (TP), falsos negativos (FN) y falsos positivos (FP) detectados, obtenidos al variar el umbral de velocidad de flujo.

Umbral de v. de flujo	Entrenamiento			Validación		
	TP	FN	FP	TP	FN	FP
1	27	17	372	13	9	197
10	28	16	293	12	10	151
20	28	16	240	14	8	132
30	28	16	196	11	11	104
40	27	17	172	11	11	87
50	27	17	147	11	11	78

4.4.3. Flujo estimado

Otro umbral es el que debe superar el flujo estimado por el filtro de Kalman de máxima correntropía para considerar el píxel en estudio como uno que pasará a la siguiente etapa. En este experimento se varía el umbral del flujo utilizando 3 valores, el propuesto en trabajos anteriores (200), 175 y 150. Este valor no puede ser demasiado bajo dado que se estaría considerando píxeles ruidosos.

En la Tabla 4.3 se muestran los resultados donde se aprecia que al disminuir el umbral, existe un leve aumento en la cantidad de falsos y una disminución de la cantidad de verdaderos positivos en el conjunto de validación. Al igual que para el caso de variar el umbral de velocidad de flujo, la disminución de los verdaderos positivos se debe a que esta corresponde a una condición del filtro individual, el cual al descartar una menor cantidad de píxeles provoca que la etapa de agrupamiento se consideren píxeles ruidosos lo que finalmente altera la detección de candidatos reales.

Tabla 4.3: Número de verdaderos positivos (TP), falsos negativos (FN) y falsos positivos (FP) detectados, obtenidos al variar el umbral de flujo.

Umbral de flujo	Entrenamiento			Validación		
	TP	FN	FP	TP	FN	FP
200	27	17	372	14	8	132
175	28	16	293	13	9	151
150	28	16	240	12	10	151

A partir de ahora, se utilizará como umbral de flujo el valor 200.

4.4.4. ASMR

Los resultados de variar este parámetro se muestran en la Tabla 4.4 en la cual se observa que a partir de un valor de 10 la cantidad de TP aumenta en una unidad (conjunto de validación) y la cantidad de FP también lo hace. Además, dado que a partir de este valor la cantidad de supernovas redescubiertas en el conjunto de validación se mantiene constante y la cantidad de FP también aumenta, se trabajará con éste.

Es importante destacar que este parámetro corresponde a una condición encontrada en la última etapa de filtrado de la rutina original, es decir, forma parte de los filtros grupales, el cual a diferencia de las condiciones del filtro de píxeles individuales no disminuye la cantidad de TP a pesar de realizar el relajamiento, dado que este solo se encargará de descartar los grupos de píxeles que no cumplan con la condición pero sin alterar su morfología.

Tabla 4.4: Número de verdaderos positivos (TP), falsos negativos (FN) y falsos positivos (FP) detectados, obtenidos al variar el parámetro ASMR.

ASMR	Entrenamiento			Validación		
	TP	FN	FP	TP	FN	FP
3	28	16	240	13	9	132
10	30	14	321	14	8	167
20	30	14	348	14	8	174
30	31	13	357	14	8	187

4.4.5. Varianza de flujo

Este parámetro también está asociado a una condición de la última etapa de filtrado, filtros grupales, por lo que un aumento de este valor también provocará un aumento en la cantidad de candidatos. En este experimento se aumentó el valor del umbral de la varianza de flujo al doble del que lo tenían los autores anteriores, donde no se observaron cambios significativos, pero dado que la idea principal es trabajar con valores más relajados, se mantendrá un valor de 500, es decir, el doble del valor de la rutina original. Los resultados de este experimento se observan en la Tabla B.3 del Apéndice B.1.

4.4.6. Modelo relajado

Un resumen de los parámetros que fueron modificados se encuentran en la Tabla 4.5, donde se comparan los valores originales con los nuevos. Además, en la Tabla 4.6 se aprecia que la cantidad de TP y FP aumentan considerablemente en los 3 conjuntos. A partir de lo anterior, se puede establecer que el modelo original es capaz de detectar una mayor cantidad de supernovas pero a un costo de obtener una mayor cantidad de falsas alarmas.

Tabla 4.5: Resumen que muestra los parámetros modificados contrastando los valores del modelo original con los del modelo relajado.

Parámetro	Valor original	Valor nuevo
neg	4	1
meg	3	1
Umbral de velocidad de flujo	50 [ADU/días]	20 [ADU/días]
Umbral de flujo	200 [ADU]	200 [ADU]
ASMR	3	10
Varianza de flujo	250 [ADU ²]	500 [ADU ²]

En la Figura 4.6 se observa un gráfico de barras que muestra la cantidad de veces que fueron detectadas las supernovas pertenecientes al conjunto de prueba, donde claramente el

Tabla 4.6: Comparación de cantidad de TP, FN y FP obtenidos al utilizar el modelo original y el modelo relajado.

Modelo	Entrenamiento			Validación			Prueba		
	TP	FN	FP	TP	FN	FP	TP	FN	FP
Original	19	25	11	9	13	19	8	15	10
Relajado	30	14	341	14	8	176	15	8	219

modelo relajado además de encontrar una mayor cantidad de supernovas las detecta más veces. En promedio las 8 supernovas en común son detectadas 1,75 veces con una desviación estándar de 0,97 para el caso del modelo original, mientras que el promedio para el caso relajado (de las mismas supernovas) es de 4,375 con una desviación estándar 2,64. Además, el promedio del número de detecciones por supernova considerando todas las encontradas en el conjunto de prueba es de 3,6 con una desviación estándar de 2,3. Lo anterior significa que el modelo relajado es capaz de seguir el crecimiento de la supernova de forma más consistente.

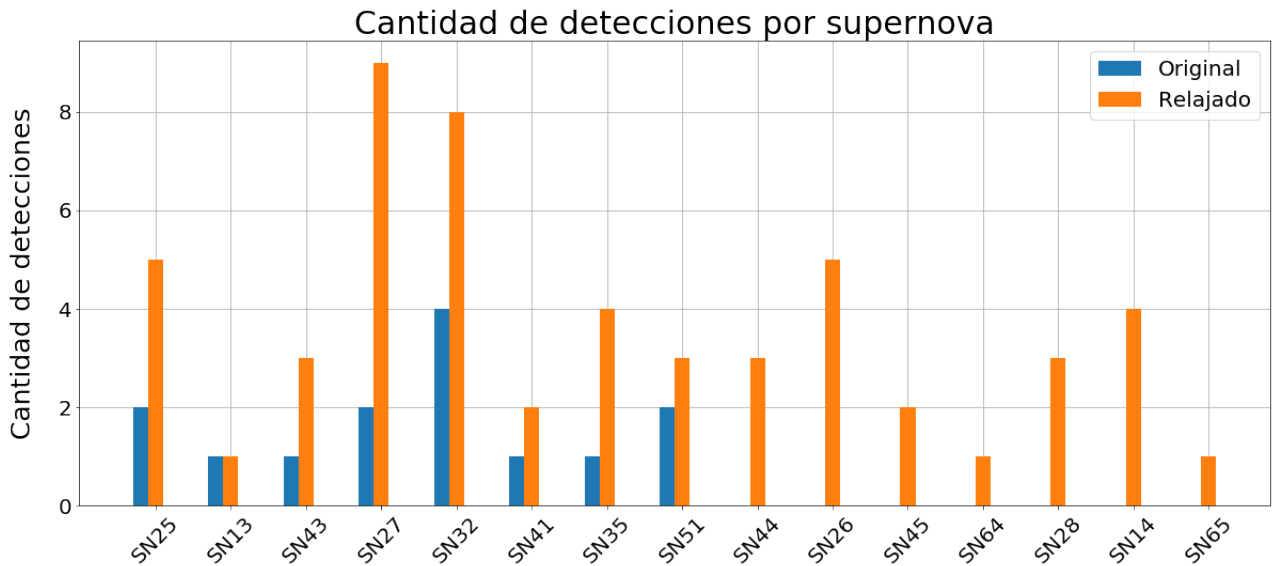


Figura 4.6: Cantidad de detecciones por supernova descubierta en el conjunto de prueba, comparando modelo original con el modelo relajado.

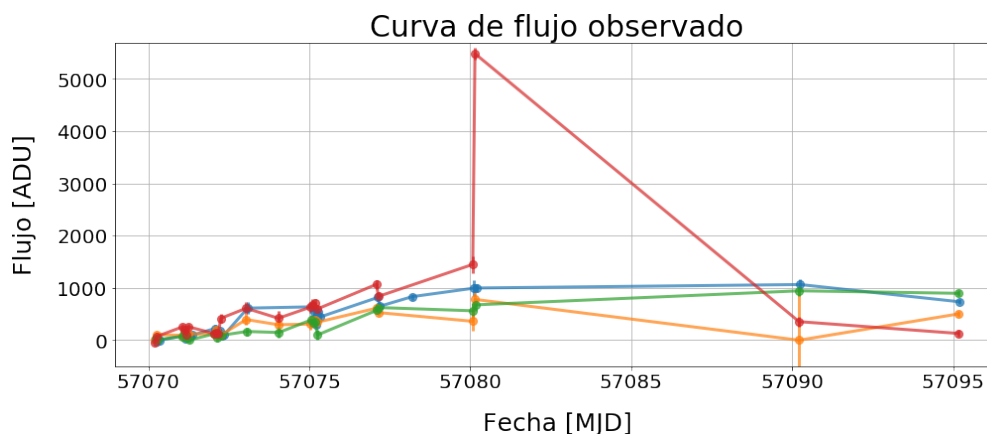
Por otro lado, en la Tabla 4.7 se muestra la diferencia de tiempo (en MJD) de detección de ambos modelos comparados con HiTS de cada supernova detectada en el conjunto de prueba. En ésta se aprecia que el modelo relajado es capaz de detectar las supernovas antes que el modelo original e incluso, antes que HiTS. Mientras el modelo original detecta las supernovas en promedio 1,14 días después que HiTS con una desviación estándar de 3,8 días mientras que el modelo relajado detecta estas mismas 8 supernovas en promedio 1,5 días antes que HiTS con una desviación estándar de 1,96 días. Además, el promedio de este último modelo considerando todas sus detecciones (en el conjunto de prueba) es de 1,65 días antes que HiTS con una desviación estándar de 2,8 días.

En la Figura 4.7 se muestran curvas de luz de algunas de las supernovas no descubiertas por el modelo original pero sí por el modelo relajado. En la Figura 4.7a se muestran las curvas obtenidas a través de métodos clásicos de fotometría (fotometría óptima de Naylor) y

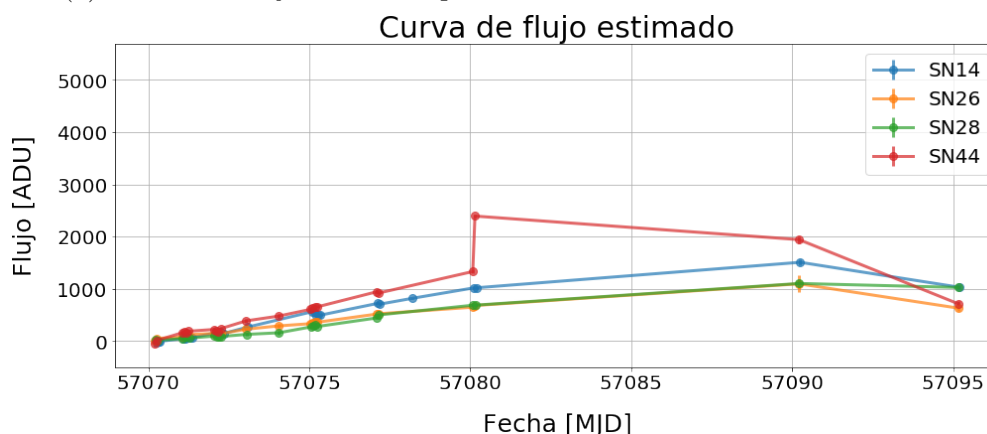
Tabla 4.7: Diferencias de tiempo (en días) con respecto a HiTS de las supernovas detectadas por la rutina original *versus* el modelo relajado. Si el valor es negativo significa que es detectado antes que la campaña HiTS, si es positivo significa que fue detectado después que la campaña HiTS, si es 0 significa que fue descubierto al mismo tiempo y es **x** si no fue descubierto por el modelo respectivo.

	Identificador de supernovas															
Modelo	25	13	43	27	32	41	35	51	44	26	45	64	28	14	65	
Original	2,0	2,0	10,0	0,0	-2,0	-3,1	1,0	-1,0	x	x	x	x	x	x	x	
Relajado	0,0	2,0	-3,0	-3,0	-4,1	-1,1	0,0	-2,9	-6,0	-2,0	-7,9	0,0	-3,0	0,1	0,0	

en la Figura 4.7b se muestran las mismas supernovas pero corregidas por el filtro de Kalman de máxima correntropía. Las curvas de flujo estimadas presentan un claro comportamiento creciente, es decir, el modelo original no es capaz de detectar todas las supernovas jóvenes (en la fase de crecimiento de su flujo). Lo anterior puede darse porque los parámetros utilizados por el modelo original limitan su detección.



(a) Curvas de flujo obtenidas por métodos tradicionales de fotometría.



(b) Curvas de flujo obtenidas por corrección de las curvas utilizando filtro de Kalman de máxima correntropía.

Figura 4.7: Curvas de flujo obtenidas de supernovas que no fueron encontradas por el modelo original pero sí por el modelo relajado. (a) curva de flujo observado, (b) curva de flujo estimado.

4.5. Entropía de curvas

A continuación se realiza un estudio en el que se evalúa la influencia de agregar un umbral de entropía a las curvas de espacio de estado y otro relacionado a la cantidad de puntos necesarios para calcular el *convex hull*, elemento clave para el cálculo de la entropía.

4.5.1. Umbral de entropía

Para comparar los resultados, se toma como base la cantidad de objetos encontrados por el modelo relajado, los cuales serán filtrados a partir de la condición de entropía de curvas.

Los resultados obtenidos al variar el valor de entropía de 0,1 a 1,0 se muestran en la Tabla 4.8, donde es posible observar que la implementación de este umbral cumple con lo esperado dado que es capaz de filtrar una gran cantidad de falsos positivos, manteniendo una gran cantidad de detecciones verdaderas. En esta tabla se observa que a menor valor de umbral de entropía, la cantidad de falsos positivos disminuyen de forma considerable, mientras que la cantidad de verdaderos positivos disminuyen en menor medida. Para umbrales de entropía menores a 0,3 la cantidad de verdaderos positivos y falsos positivos son considerablemente más bajos, lo cual no es conveniente dado que se pierde una gran cantidad de supernovas. Además, es posible observar que incluso utilizando valores relativamente altos del umbral de entropía es posible discriminar una gran cantidad de falsos positivos. Esto último indica que el estudio de la curva en el espacio de estados sí aporta información para discriminar candidatos.

Tabla 4.8: Número de verdaderos positivos (TP), falsos negativos (FN) y falsos positivos (FP) detectados, obtenidos al variar el umbral de entropía utilizado como base el modelo relajado.

Umbral de Entropía	Entrenamiento			Validación		
	TP	FN	FP	TP	FN	FP
-	30	14	341	14	8	176
1	29	15	195	13	9	120
0,9	29	15	152	13	9	105
0,8	29	15	123	12	10	91
0,7	29	15	92	12	10	79
0,6	29	15	76	12	10	71
0,5	28	16	66	12	10	57
0,4	28	16	56	11	11	42
0,3	26	18	44	11	11	21
0,2	23	21	32	8	14	8
0,1	17	27	7	6	16	7

En la Figura 4.8 se observa un gráfico *recall vs precision* de los resultados obtenidos de los modelos relajados con umbral de entropía (en el gráfico aparece el valor del umbral de entropía correspondiente), del modelo relajado sin umbral de entropía (en el gráfico aparece como “sin”) y el modelo original (en el gráfico aparece como “original”).

El primer criterio para determinar qué valores seguirán siendo estudiados es que la solución

posea un mayor valor de *recall* que el caso original, dado que esto significa que es posible encontrar una mayor proporción de eventos verdaderos. Además, el modelo escogido debe poseer una alta *precision*, dado que esto indicaría que la proporción de detecciones correctas reportadas por éste es alta. En este caso en particular, el punto que está a la derecha y arriba del punto correspondiente al modelo original, en el conjunto de validación, es el punto asociado a un umbral de entropía 0,3. A pesar que este valor en particular pareciera ser uno de los mejores, aún es interesante estudiar otros modelos que posean un alto *recall* dado que estos detectan una mayor cantidad de supernovas. Se espera que a través del ajuste paramétrico de elementos del filtro de Kalman de máxima correntropía se puedan filtrar aún más falsos positivos, manteniendo la mayor cantidad de verdaderos positivos.

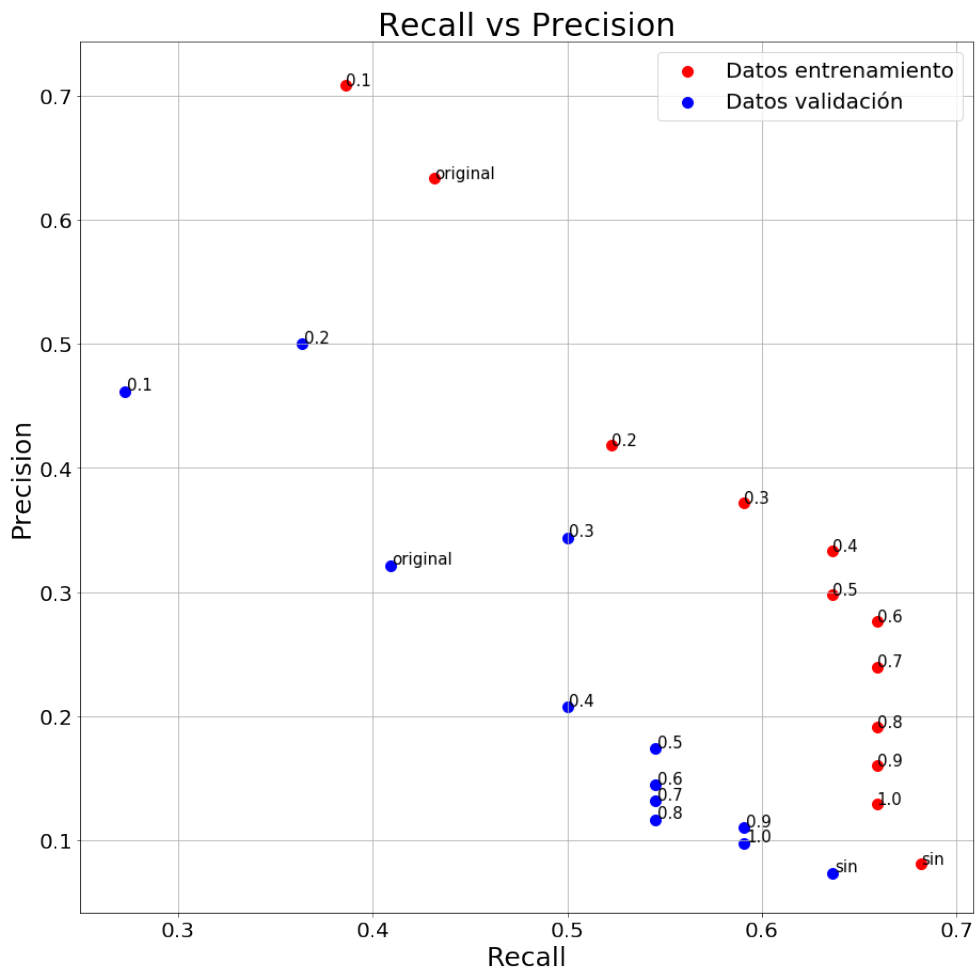


Figura 4.8: Gráfico *recall vs precision* ante distintos modelos en los que se varía solo el umbral de entropía. Se incluye además los resultados de autores anteriores, los cuales corresponden a los puntos marcados con la palabra “original”.

En las futuras secciones se realizarán pruebas utilizando los modelos cuyos umbrales de entropía son 0,3 y 0,6. Estos valores son escogidos por poseer un buen *trade off* entre la cantidad de verdaderos y falsos positivos, además de poseer valores de *recall* y *precision* que indican una buena proporción de detecciones positivas.

4.5.2. Cantidad de puntos para calcular entropía

Otro de los parámetros importantes que agrega esta nueva condición, es la cantidad de puntos que se utilizan para determinar la entropía de una curva (en particular, para realizar el cálculo del *convex hull*), donde se espera encontrar una cantidad que permita discriminar aún más falsos positivos, sin la necesidad de usar toda la información recopilada. A través de las pruebas realizadas, al variar la cantidad de puntos que son considerados al calcular la entropía de curvas, la cantidad de detecciones correctas (TP) y detecciones falsas (FP) se ven afectadas. Al considerar una baja cantidad de estos puntos, la discriminación de supernovas es baja en comparación al usar una mayor cantidad de puntos, puesto que esta última es capaz de considerar más información. Por otro lado, no es necesario utilizar todos los puntos de los que se posea información, dado que a partir de cierto valor (en este caso, 16), es posible obtener la misma cantidad de TP Y FP. Lo anterior puede ser corroborado al observar la Tabla B.4 ubicada en el Apéndice B.1.

4.5.3. Modelos relajados con umbrales de entropía

En la Tabla 4.9 se observa una comparación de la cantidad de detecciones producidas por el modelo original y tres modelos relajados con distintos umbrales de entropía. Se observa claramente que los modelos con umbrales poseen una mayor cantidad de detecciones positivas pero a su vez siguen teniendo más falsos positivos que el modelo original, esto se aprecia claramente en el modelo relajado con umbral 0,9. A pesar de esto último, los resultados obtenidos por el caso de entropía 0,3 poseen un número manejable de falsos positivos que se acercan a las cantidades del caso original. Al realizar la comparación entre los tres modelos con umbrales, el que posee un valor de 0,9 es capaz de detectar un par más de supernovas (en el caso del conjunto de prueba) pero con una gran cantidad de falsos positivos. Además, al realizar una comparación entre el modelo relajado que se observa en la Tabla 4.6 y cualquier modelo relajado que considere un umbral de entropía, se observa una clara disminución en la cantidad de falsos positivos y una leve disminución en la cantidad de verdaderos positivos.

Tabla 4.9: Número de verdaderos positivos (TP), falsos negativos (FN) y falsos positivos (FP) detectados por modelo original y modelos relajados con distintos umbrales de entropía.

Modelo	Umbral de entropía	Entrenamiento			Validación			Prueba		
		TP	FN	FP	TP	FN	FP	TP	FN	FP
Original	-	19	25	11	9	13	19	8	15	10
Relajado	0,3	26	18	44	11	11	21	12	11	22
Relajado	0,5	28	16	66	12	10	57	14	9	33
Relajado	0,9	29	15	152	13	9	105	15	8	124

En la Figura 4.9 se observan 4 modelos distintos, 3 modelos relajados con distinto umbral de entropía y el modelo original. En general los 3 modelos relajados detectan una mayor cantidad de supernovas que el modelo original y además éstas son detectadas una mayor cantidad de veces. Esto significa que a pesar de que al modelo relajado se le agrega un umbral de entropía, sus beneficios (mayor cantidad de supernovas detectadas y mayor cantidad de detecciones por supernova) se mantienen.

Las pequeñas diferencias observables en la Figura 4.9 son para el caso SN27, donde el modelo que posee un mayor umbral (es decir, filtra una menor cantidad de datos) posee más detecciones de la supernova que el resto de los casos. A pesar de lo anterior, los 3 casos logran mejorar la cantidad de detecciones en comparación al original.

Por otro lado, en la Tabla 4.10, se observa que para los 3 modelos con umbrales de entropía, existe una mayor cantidad de detecciones y ocurren, en promedio, antes que el modelo original e incluso que HiTS.

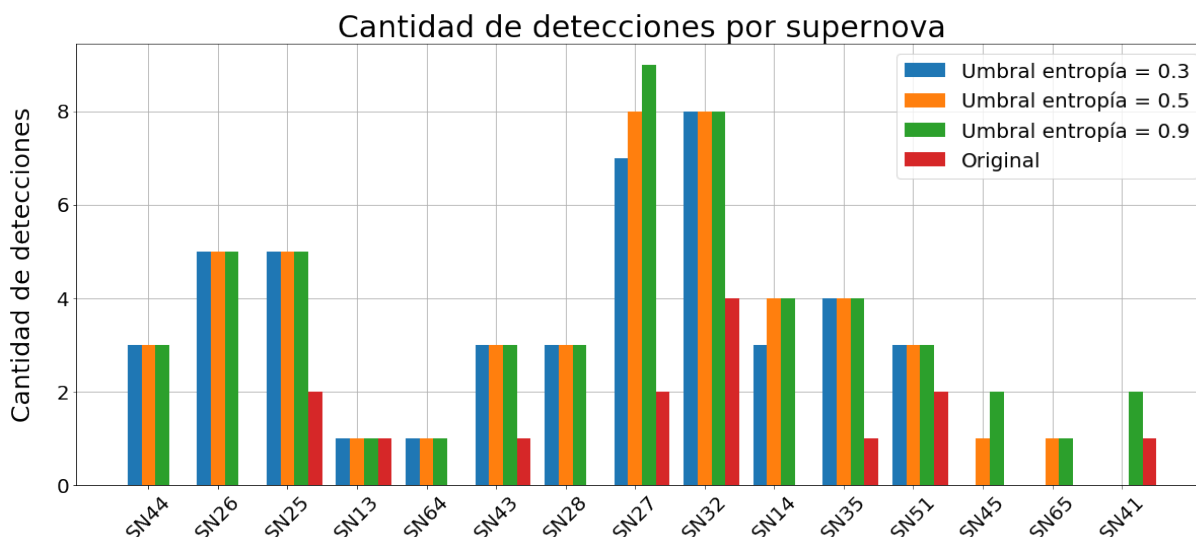


Figura 4.9: Cantidad de detecciones por supernova descubierta en el conjunto de prueba, comparando el modelo original con los modelos relajados con distinto umbral de entropía.

Tabla 4.10: Diferencias de tiempo (en días) con respecto a HiTS de las supernovas detectadas por el modelo original *versus* los modelos relajados con distintos umbrales de entropía. Si el valor es negativo significa que es detectado antes que la campaña HiTS, si es positivo significa que fue detectado después que la campaña HiTS, si es 0 significa que fue descubierto al mismo tiempo y es **x** si no fue descubierto por el determinado modelo.

Modelo	Umbral de Entropía	Identificador de supernovas														
		25	13	43	27	32	41	35	51	44	26	45	64	28	14	65
Original	-	2,1	2,0	10,1	0,1	-2,0	-3,1	1,0	-1,1	x	x	x	x	x	x	x
Relajado	0,3	0,1	2,0	-3,0	-3,0	-4,1	x	0,0	-2,9	-6,0	-2,0	0,0	3,0	2,0	x	x
Relajado	0,5	0,1	2,0	-3,0	-3,0	-4,1	x	0,0	-2,9	-6,0	-2,0	0,0	3,0	0,1	-7,9	0,0
Relajado	0,9	0,1	2,0	-3,0	-3,0	-4,1	-1,1	0,0	-2,9	-6,0	-2,0	0,0	3,0	0,1	-7,9	0,0

4.6. Filtro de Kalman de máxima correntropía

4.6.1. Parámetro sigma del modelo

El siguiente experimento se realizó utilizando el **modelo relajado con umbral de entropía 0,6** dado que este posee una mayor cantidad de candidatos que el modelo con umbral de entropía 0,3, lo cual permite observar con mayor claridad si la modificación del σ_{modelo}

permite reducir la cantidad de falsos positivos. En la Tabla 4.11 se observa que a medida que aumenta el valor de σ_{modelo} la cantidad de falsos positivos disminuye. Lo anterior se asocia a la diferencia de tiempo existente entre observaciones, la cual si es muy alta provocará que el filtro de Kalman de máxima correntropía se acerque más al valor de la observación (con respecto a valores más pequeños del parámetro) que a la del modelo, dado que el error de este último se vería aumentado. Una representación de lo anterior se observa en la Figura 4.10, donde se observa que el punto estimado ubicado en la fecha 57090 [MJD] sufre una considerable disminución de su valor de flujo y su punto predicho aumenta considerablemente su error al pasar a un $\sigma_{modelo} = 0,4$ por su gran diferencia temporal entre este y el punto anterior. Además, se observa que a partir de este valor ($\sigma_{modelo} = 0,4$) la diferencia de falsos positivos es prácticamente nula, es por esto que se mantendrá este valor para el desarrollo de siguientes experimentos.

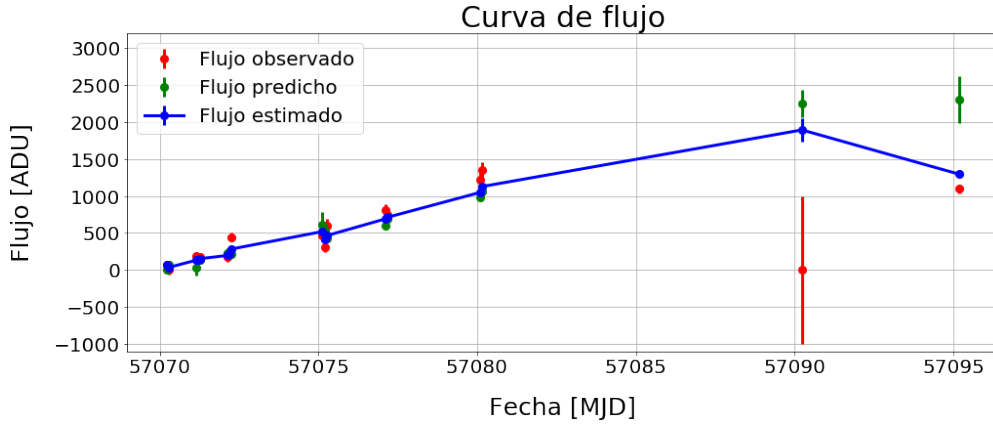
Dado que el objetivo es alcanzar una baja cantidad de falsos positivos, en los siguientes experimentos se trabajará con el modelo relajado que posee un umbral de entropía de 0,3 y un $\sigma_{modelo} = 0,4$ puesto que con este se logró obtener la misma cantidad de TP que para el modelo con umbral de entropía de 0,6 en el conjunto de validación y además se redujo la cantidad de falsos positivos considerablemente. Lo anterior se puede apreciar en la Tabla B.5 que se encuentra en el Apéndice B.1.

Tabla 4.11: Número de verdaderos positivos (TP), falsos negativos (FN) y falsos positivos (FP) detectados, obtenidos al variar el parámetro sigma del modelo matemático. Se utilizó como base el modelo relajado con umbral de entropía 0,6.

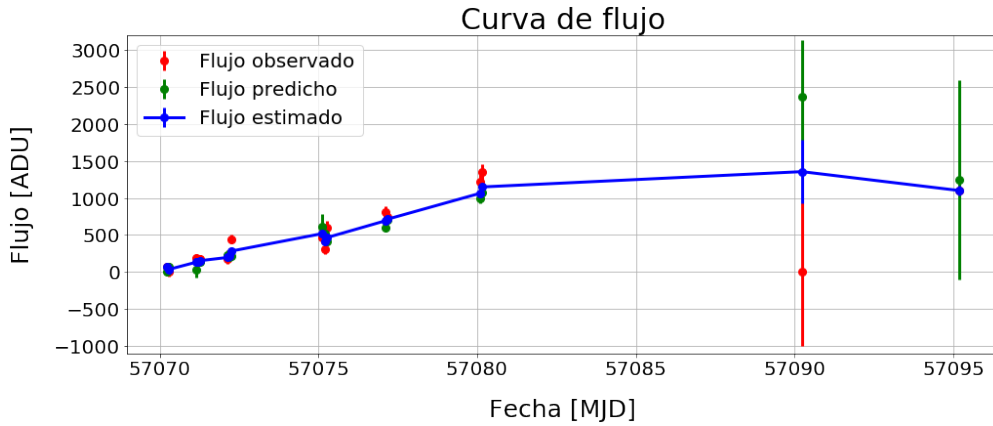
σ_{modelo}	Entrenamiento			Validación		
	TP	FN	FP	TP	FN	FP
0,1	29	15	70	12	10	62
0,2	27	17	47	11	11	60
0,3	26	18	44	11	11	55
0,4	26	18	44	11	11	53
0,5	26	18	45	11	11	52

4.6.2. Ancho de banda de *kernel* Gaussiano

En la Tabla 4.12 se muestran los resultados de variar el valor de $\sigma_{correntropia}$ en el modelo relajado con umbral de entropía 0,3, el cual representa el ancho de banda del *kernel* Gaussiano y es utilizado para calcular la correntropía. En ésta se observa que no existen diferencias en la cantidad de TP detectados en el conjunto de validación para los primeros 3 valores (1000, 400, 200) y que en validación solo existen 2 FP de diferencia entre el modelo 1000 y los modelos con valores de $\sigma_{correntropia}$ 400 y 200. Por otro lado, al utilizar valores muy bajos (por ejemplo 10) se observa que la cantidad de TP se ve disminuida y la cantidad de FP aumenta incluso más que para el caso de $\sigma_{correntropia} = 1000$, puesto que si este valor es pequeño, el algoritmo tiende a un impulso de Dirac, lo que provoca que diverja o que converja muy lentamente [10]. Para este caso en particular se iteró máximo 10 veces, lo cual no fue suficiente para que el algoritmo alcance a converger. Dado que para los valores de $\sigma_{correntropia}$ 400 y 200 no existen diferencia en la cantidad de TP, FN y FP de su conjunto de validación, en teoría se podrían escoger los 2, pero como el valor 200 es más pequeño, éste permitiría eliminar con mayor facilidad los *outliers* que llegasen a aparecer, es por esto que se escogerá $\sigma_{correntropia} = 200$.



(a) Curva de flujo utilizando $\sigma_{modelo} = 0, 1$



(b) Curva de flujo utilizando $\sigma_{modelo} = 0, 4$

Figura 4.10: Curvas de flujo estimadas por el filtro de Kalman de máxima correntropía, predichas y observadas de supernova ubicada en el campo 27, CCD 10, perteneciente al conjunto de entrenamiento. En (a) se observa la curva de flujo utilizando $\sigma_{modelo} = 0, 1$, mientras que en (b) se observa esta misma curva, utilizando $\sigma_{modelo} = 0, 4$.

Dado que los resultados obtenidos por el modelo relajado con umbral de entropía 0,3 poseen una baja cantidad de candidatos, se realizaron pruebas utilizando el modelo relajado sin aplicar la condición del umbral de entropía. Al realizar lo anterior se observó el mismo comportamiento que utilizando el modelo anterior. Lo anterior se observa en la Tabla B.6 en el Apéndice B.1.

Tabla 4.12: Número de verdaderos positivos (TP), falsos negativos (FN) y falsos positivos (FP) detectados, obtenidos al variar $\sigma_{correntropia}$, usando como modelo base el modelo relajado con umbral de entropía 0,3.

$\sigma_{correntropia}$	Entrenamiento			Validación		
	TP	FN	FP	TP	FN	FP
1000	24	20	18	11	11	18
400	24	20	20	11	11	16
200	24	20	13	11	11	16
10	16	28	115	8	14	64

4.6.3. Modelo final (sin poda)

En la Figura 4.11 se observa la cantidad de detecciones de supernovas encontradas en el conjunto de prueba. Se aprecia que el parámetro $\sigma_{correntropia}$ no altera significativamente la cantidad de detecciones por supernova en los modelos que utilizan $\sigma_{correntropia} = 1000$ y $\sigma_{correntropia} = 200$, mientras que para un $\sigma_{correntropia} = 10$ existen diferencias notorias, donde se dejan de encontrar algunas supernovas que sí detecta el modelo original (SN13, SN51, SN41 y SN43). Los modelos con $\sigma_{correntropia}$ iguales a 1000 y 200 logran detectar más veces las supernovas que descubren en común con el modelo original, además de poder encontrar un mayor número. Solo existe un caso (SN43) que el modelo original es capaz de encontrarlo.

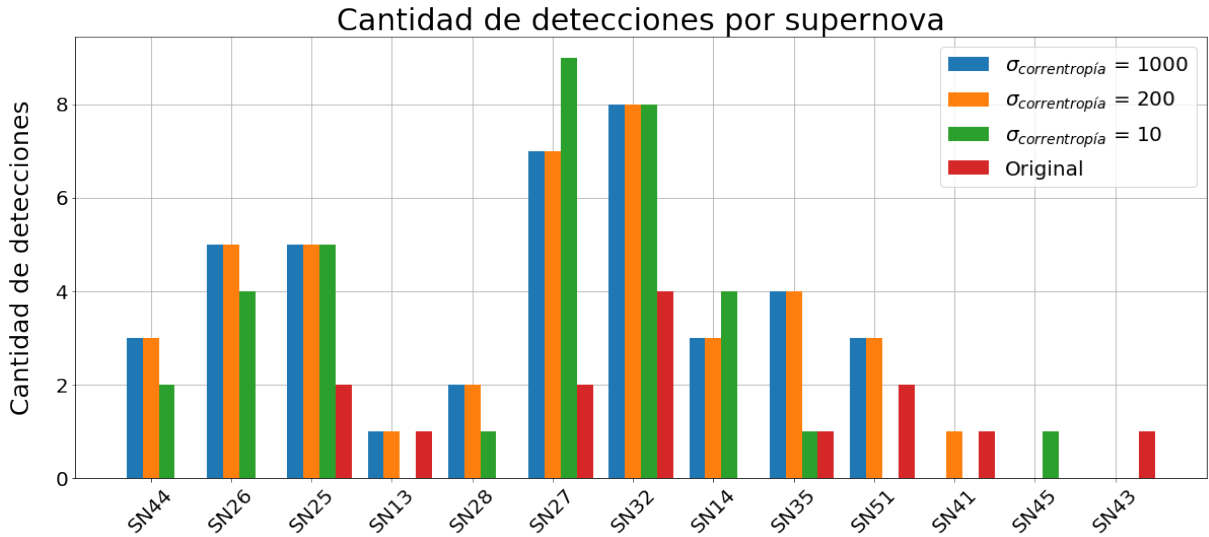


Figura 4.11: Cantidad de detecciones por supernova descubierta en el conjunto de prueba. Se compara modelo original con modelos que utilizan distintos umbrales de entropía.

En la Tabla 4.13 se observa nuevamente que para la mayoría de los casos (excepto SN43 y SN41) los modelos con $\sigma_{correntropia}$ igual a 1000 y 200 son capaces de encontrar las supernovas antes que el modelo original y en muchos casos antes que HiTS. Las supernovas descubiertas tanto por el modelo con un valor de 1000 y 200 son detectadas al mismo tiempo (excepto SN28).

Tabla 4.13: Diferencias de tiempo (en días) con respecto a HiTS de las supernovas detectadas por el modelo original *versus* los modelos relajados con umbral de entropía 0,3, σ_{modelo} 0,4 y distintos valores de $\sigma_{correntropia}$. Si el valor es negativo significa que es detectado antes que la campaña HiTS, si es positivo significa que fue detectado después que la campaña HiTS, si es 0 significa que fue descubierto al mismo tiempo y es x si no fue descubierto por el determinado modelo.

Modelo	σ_{modelo}	Umbral de Entropía	$\sigma_{correntropia}$	Identificador de supernovas															
				25	13	43	27	32	41	35	51	44	26	45	64	28	14	65	
Original	0,1	-	1000	2,1	2,0	10,1	0,1	-2,0	-3,1	1,0	-1,1	x	x	x	x	x	x		
Relajado	0,4	0,3	1000	0,1	2,0	x	-3,0	-4,1	x	0,0	-2,9	-6,0	-2,0	3,0	2,0	2,0	x	x	
Relajado	0,4	0,3	200	0,1	2,0	x	-3,0	-4,1	2,0	0,0	-2,9	-6,0	-2,0	3,0	2,0	0,1	-7,9	0,0	
Relajado	0,4	0,3	10	0,0	2,0	-3,0	-4,8	-4,1	2,0	0,0	-2,9	-7,0	-2,0	3,0	2,0	0,	-7,9	0,0	

Finalmente, se escogerá un valor de $\sigma_{correntropia}$ igual a 200 puesto fue el que mejores resultados presentó.

El modelo generado hasta este punto, cuyos parámetros se muestran en la Tabla 4.14, será considerado como el **modelo final** al cual se le realizarán otras pruebas con el fin de tener un análisis más completo de éste.

Tabla 4.14: Resumen de los parámetros y metaparámetros utilizados por el modelo original y modelo final.

Modelo	Parámetros relajados	<i>Zero points</i>	$\sigma_{correntropia}$	σ_{modelo}	Umbral de Entropía	Lim
Final	Si	Si	200	0,4	0,3	16
Original	No	No	1000	0,1	-	-

Una vez relajado, optimizado y estudiado los parámetros de interés, además de incorporar el cálculo de entropía de curvas, se procede a realizar una poda, es decir, se eliminarán condiciones que al quitarlas no alteren considerablemente el funcionamiento del programa.

4.7. Poda

Dado que es costoso computacionalmente escoger una combinación óptima para eliminar las condiciones, éstas serán eliminadas progresivamente hasta que exista una diferencia relevante en los resultados.

En este experimento se eliminaron desde 0 hasta 7 condiciones, donde no se observaron cambios relevantes entre sí. De hecho, al eliminar hasta 3 condiciones los resultados en los conjuntos de entrenamiento y de validación no presentan cambio alguno. Al seguir eliminando condiciones la cantidad de TP se mantienen iguales en ambos conjuntos y la cantidad de FP aumenta levemente. Con este experimento se mostró que al eliminar 7 condiciones los resultados prácticamente no varían. Lo anterior puede observarse en la Tabla B.7 en el Apéndice B.1.

Las condiciones eliminadas progresivamente son las que se muestran a continuación:

1. Si la varianza del flujo estimado por el filtro de Kalman de máxima correntropía del píxel en estudio es superior al umbral impuesto (umbral = 150 [ADU/días]).
2. Si el píxel central es rechazado al ser superior a la mediana.
3. Si el píxel central del grupo se encuentra dentro de los píxeles considerados como defectuosos (información entregada por máscara).
4. Si no existen máximos locales alrededor de la posición del píxel central en estudio en la imagen científica.
5. Si la varianza del píxel central del grupo es mayor al determinado por el umbral. Filtro de píxeles individuales.
6. Si el píxel no ha sido descartado por superar la mediana.
7. Si el grupo posee algún píxel que doble el valor del flujo o imagen científica.

8. Si no hay máximos locales dentro del grupo encontrado en la matriz de flujo.

Las condiciones 1 y 4 pertenecen al filtrado de píxeles individuales mientras que el resto pertenece al filtro por grupos.

4.7.1. Modelo final podado

En la Tabla 4.15 se aprecia un resumen del modelo original y el comportamiento del modelo final con distinta cantidad de condiciones eliminadas. En primer lugar, se observa que en los 3 conjuntos (entrenamiento, validación y prueba) la cantidad de TP de los modelos finales es superior a la cantidad del modelo original. Además, el modelo final con 7 condiciones eliminadas resulta ser el que posee una mayor cantidad de TP y FP (en su conjunto de prueba y entrenamiento) dado que este es más flexible que los otros modelos pues no debe satisfacer tantas reglas. Esto indica que a pesar de que se eliminan condiciones, el comportamiento del detector de supernova no sufre variaciones considerables e incluso es capaz de detectar más supernovas confirmadas por HiTS.

Tabla 4.15: Número de verdaderos positivos (TP), falsos negativos (FN) y falsos positivos (FP) detectados por modelo original y modelos finales con distintos números de condiciones eliminadas.

Modelo	N° de condiciones eliminadas	Entrenamiento			Validación			Prueba		
		TP	FN	FP	TP	FN	FP	TP	FN	FP
Original	0	19	25	11	9	13	19	8	15	10
Final	0	24	20	13	11	11	16	11	12	10
Final	3	24	20	13	11	11	16	11	12	11
Final	7	25	19	29	11	11	18	12	11	18

En la Figura 4.12 se aprecia que el modelo con 7 condiciones eliminadas es capaz de detectar una supernova más (SN14). Además, se observa que no existe una diferencia considerable en la cantidad de veces que son detectadas las supernovas del conjunto de prueba entre los modelos con número de condiciones eliminadas 0, 3 y 7, solo la supernova SN44 presenta una mayor cantidad de detecciones por parte del modelo con 7 condiciones eliminadas. Por otro lado, el modelo original es el que detecta menor cantidad de veces las supernovas detectadas del conjunto de prueba.

En la Tabla 4.16 se observa que los modelos finales con una distinta cantidad de condiciones eliminadas, logran encontrar más supernovas, en promedio, antes que el modelo original y en muchos casos, antes que HiTS. Además, entre los modelos finales solo existen diferencias en la supernova SN43, que solo el modelo con una mayor cantidad de condiciones eliminadas logra detectar y además en la supernova SN44 que se logra detectar un día antes que el resto de los modelos finales.

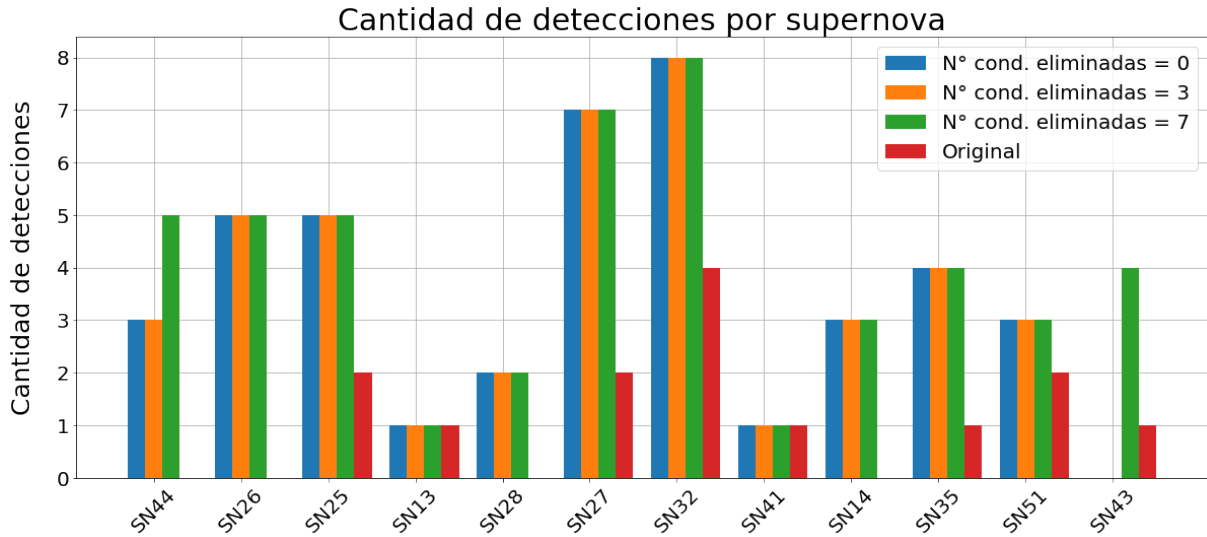


Figura 4.12: Cantidad de detecciones por supernova descubierta en el conjunto de prueba, comparando modelo original con modelo finales con distintas cantidades de condiciones eliminadas.

Tabla 4.16: Diferencias de tiempo (en días) con respecto a HiTS de las supernovas detectadas por el modelo original *versus* modelos finales con distintas cantidades de condiciones eliminadas. Si el valor es negativo significa que es detectado antes que la campaña HiTS, si es positivo significa que fue detectado después que la campaña HiTS, si es 0 significa que fue descubierto al mismo tiempo y es **x** si no fue descubierto por el determinado modelo.

Modelo	Condiciones Eliminadas	Identificador de supernovas										
		25	13	43	27	32	41	35	51	44	26	28
Original	0	2,1	2,0	10,1	0,1	-2,0	-3,1	1,0	-1,1	x	x	x
Final	0	0,1	2,0	x	-3,0	-4,1	2,0	0,0	-2,9	-6,0	-2,0	3,0
Final	3	0,1,	2,0	x	-3,0	-4,1	2,0	0,0	-2,9	-6,0	-2,0	3,0
Final	7	0,1,	2,0	-3,0	-3,0	-4,1	2,0	0,0	-2,9	-7,0	-2,0	3,0

4.8. Otras Pruebas

4.8.1. Comparación de modelos obtenidos en distintas secciones

En la Tabla 4.17 se proyecta el resumen de 4 modelos obtenidos a lo largo del presente trabajo, donde se muestra la cantidad de TP, FN y FP de los conjuntos de entrenamiento, validación y prueba. En ésta, es claro notar que para todos los casos la cantidad de TP supera a los del modelo original. Si bien es cierto, el modelo relajado es el que presenta la mayor cantidad de TP este posee una gran cantidad de FP lo cual lo hace un modelo poco deseado. Por otro lado, el modelo final logra encontrar más TP que el caso original e incluso una menor cantidad de FP para los conjuntos de validación y prueba. Además, el modelo final podado logra encontrar más supernovas que el caso del modelo final (sin poda) a un costo de encontrar una mayor, pero manejable, cantidad de falsos positivos. El beneficio de este último modelo es que al simplificar el modelo, es más fácil analizar sus parámetros y se

puede evitar sobreajustar los datos.

Tabla 4.17: Comparación de cantidad de TP, FN y FP obtenidos al utilizar los modelos: original, relajado, final y final podado.

Modelo	Entrenamiento			Validación			Prueba		
	TP	FN	FP	TP	FN	FP	TP	FN	FP
Original	19	25	11	9	13	19	8	15	10
Relajado	30	14	341	14	8	176	15	8	219
Final	24	20	13	11	11	16	11	11	10
Final podado	25	19	29	11	11	18	12	11	18

En la Figura 4.13 se observa que la cantidad de detecciones de los nuevos modelos (relajado, final y final podado) son bastante parecidos entre sí, pero el primero de estos (modelo relajado) es el que realiza una mayor cantidad de detecciones y, en su mayoría, es capaz de detectar más veces la misma supernova en comparación al resto de los modelos. Por otro lado, el modelo original no solo detecta menos supernovas, sino que también los detecta una menor cantidad de veces, es decir, no es consistente en la detección.

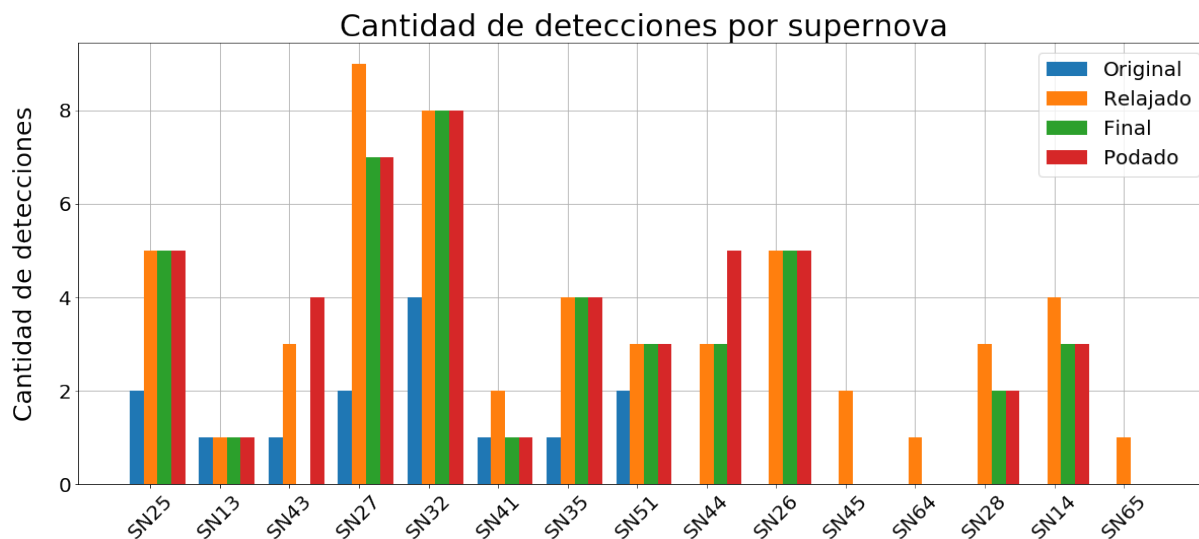


Figura 4.13: Cantidad de detecciones por supernova descubierta en el conjunto de pruebas (*test*), comparando los modelos: original, relajado, final y final podado.

En la Tabla 4.18 se observa que para todas las supernovas encontradas (a excepción de la supernova SN41) los 3 modelos propuestos son capaces de encontrarlas antes que el modelo original. Además, entre estos 3 modelos no se observan grandes diferencias en el tiempo de detección, lo cual significa que el aplicar el umbral de entropía y el variar los metaparámetros del filtro de Kalman de máxima correntropía, no altera considerablemente los resultados del modelo relajado.

En la Figura 4.14 se observan 3 series de estampillas relacionada a los píxeles considerados como partes de una supernova. Se aprecia que para el caso del modelo relajado (Figura 4.14b) existe un leve aumento en los píxeles catalogados como candidatos con respecto al modelo

Tabla 4.18: Diferencias de tiempo (en días) con respecto a HiTS de las supernovas detectadas por los modelos original, relajado, final y final podado (7 condiciones eliminadas). Si el valor es negativo significa que es detectado antes que la campaña HiTS, si es positivo significa que fue detectado después que la campaña HiTS, si es 0 significa que fue descubierto al mismo tiempo y es **x** si no fue descubierto por el determinado modelo.

Modelo	Identificador de supernovas														
	25	13	43	27	32	41	35	51	44	26	45	64	28	14	65
Original	2,1	2,0	10,1	0,1	-2,0	-3,1	1,0	-1,1	x	x	x	x	x	x	x
Relajado	0,1	2,0	-3,0	-3,0	-4,1	-1,1	0,0	-2,9	-6,0	-2,0	-7,9	0,0	3,0	0,1	0,0
Final	0,1	2,0	x	-3,0	-4,1	2,0	0,0	-2,9	-6,0	-2,0	x	x	3,0	2,0	x
Final Podado	0,1	2,0	-3,0	-3,0	-4,1	2,0	0,0	-2,9	-7,0	-2,0	x	x	3,0	2,0	x

original (Figura 4.14a). Además, en la tercera serie de imágenes (Figura 4.14c) se observa un nuevo aumento leve de la cantidad de píxeles de algunas estampillas.

Dado lo anterior, a medida que se relaja y/o podan las condiciones relacionadas a la etapa de filtrado individual de píxeles, la cantidad de píxeles candidatos aumenta, lo cual altera a la etapa de agrupamiento de píxeles, provocando cambios en el píxel central en estudio, pudiendo provocar que éste quede en una zona que no supere las condiciones necesarias para ser considerado como candidato a supernova. Esto último sería el motivo del porqué en algunos casos, al relajar las condiciones, la cantidad de TP se veían alteradas negativamente, es decir, a pesar de usar umbrales más bajos, la cantidad de TP disminuían.



Figura 4.14: Estampillas correspondientes a los píxeles candidatos de la supernova del conjunto de entrenamiento ubicada en el campo 07 y CCD N14. En (a) se encuentra la serie correspondiente al modelo original, en (b) se encuentra la serie que corresponde al modelo relajado y en (c) se encuentra la serie correspondiente al modelo final podado.

En la Figura 4.15 se observa una serie de curvas de flujo del conjunto de entrenamiento que todos los modelos fueron capaces de encontrar donde se observa un claro comportamiento creciente en cada uno de estos, lo cual es un comportamiento típico de las supernovas. En la Figura 4.16 se observan las curvas de TP que el modelo original no detectó. Es posible apreciar que, a grandes rasgos, las curvas de ambas figuras poseen un comportamiento creciente similar al esperado de una supernova. Esto demuestra que el modelo original ignoró supernovas cuyas curvas de flujo poseen comportamientos similares a las que sí logró encontrar.

Al comparar las curvas en el espacio de estados de las supernovas del conjunto de entrenamiento que fueron descubiertas por todos los modelos (ver Figura 4.17), con las que fueron

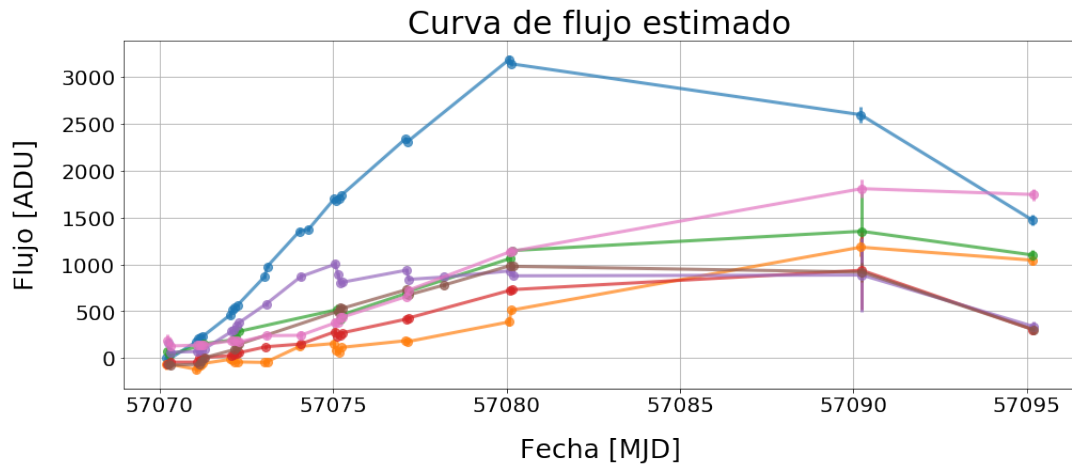


Figura 4.15: Ejemplos de curvas de flujo del conjunto de entrenamiento, descubiertas por los modelos original, relajado, final y final podado.

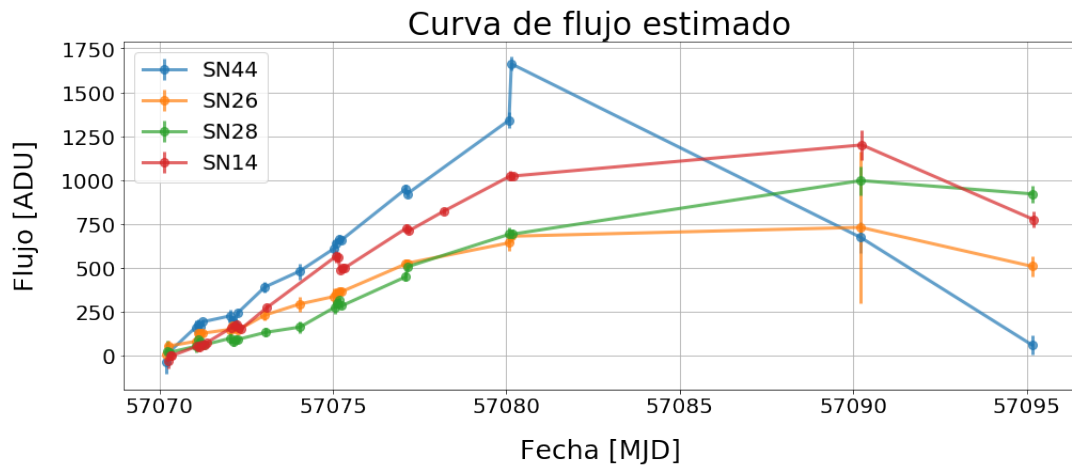


Figura 4.16: Ejemplos de curvas de flujo del conjunto de prueba, descubiertas solo por los modelos relajado, final y final podado.

descubiertas solo por los nuevos modelos (ver Figura 4.18) se observa, a grandes rasgos, una forma similar en las curvas. Al principio de la curva se aprecia un aumento tanto en el flujo como en la velocidad de flujo, luego este último decrece lentamente y finalmente los últimos 2 puntos muestran cambios abruptos en la curva, los cuales se deben a que estos se obtienen luego de un largo periodo sin observaciones, es decir, se pierde continuidad entre puntos. Además, en ambas figuras se aprecian 2 líneas rectas discontinuas que corresponden a los umbrales de flujo y de velocidad de flujo que debe superar cada punto para ser considerado como candidato.

A continuación, se describe el comportamiento de los tipos de falsos positivos que más se repiten dentro de los candidatos obtenidos por el modelo final (sin poda).

En la Figura 4.19 se observa uno de los falsos positivos descubiertos, el cual se trataría de un candidato a estrella variable debido a que su curva de luz presenta un incremento de su flujo y luego éste decrece en un periodo corto de tiempo. En la Figura 4.19a se observa

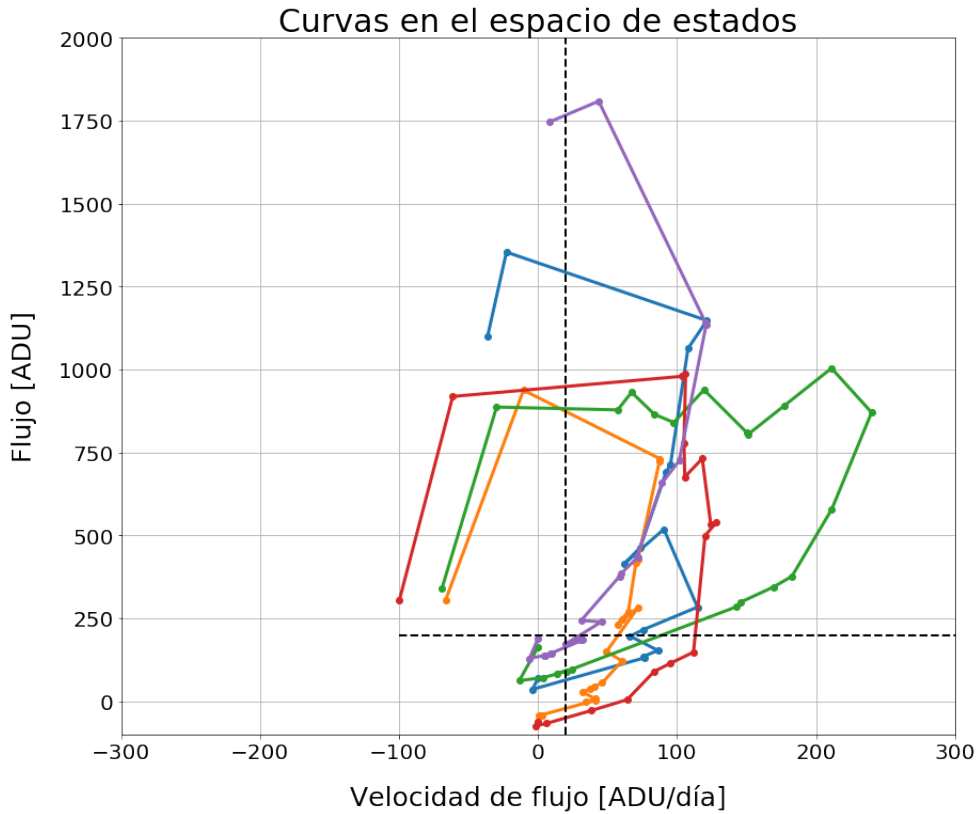


Figura 4.17: Ejemplos de curvas en espacio de estados de supernovas pertenecientes al conjunto de entrenamiento, descubiertas por todos los modelos original, relajado, final y final podado. Además, se aprecian 2 líneas rectas discontinuas de color negro, las cuales corresponden a los umbrales de flujo (200 [ADU]) y de velocidad de flujo (20 [ADU/días]).

el comportamiento descrito anteriormente, mientras que en la Figura 4.19b se observa tanto en la serie de imágenes de diferencia como en la flujo observado una intermitencia en la luminosidad del objeto, donde en las primeras imágenes pareciera tener un comportamiento de supernova (ocurre un claro incremento en su flujo) pero luego de haber disminuido su flujo pareciera volver a aumentar.

Este tipo de falsos positivos ocurre porque el comportamiento inicial del objeto es similar al de una supernova, la diferencia radica en que si al seguir observando esa zona del cielo, el objeto luego de “apagarse” (como lo haría una supernova) vuelve a aumentar su flujo de forma idéntica al inicial, se trataría de un candidato a estrella variable.

Otro tipo de falso positivo que se repite en el programa, es el denominado **crecimiento sutil**, cuyas figuras se muestran en la Figura 4.20. Es fácil apreciar que con respecto al flujo (ver Figura 4.20a) existe un leve aumento en el tiempo. Dado que la curva presenta un crecimiento leve el programa se confunde y emite alertas. A simple vista, en la Figura 4.20b no se observa ningún cambio de interés en sus imágenes científicas, de diferencia ni en el flujo observado. En las estampillas de *pixel flags* se observan píxeles que fueron capaces de pasar la primera etapa de filtrado, mientras que en las estampillas de *groups flags* se muestran los grupos de píxeles que fueron descartados (a través de los filtros grupales), en donde se aprecia que para este caso en particular no se descartaron los píxeles encontrados. Debido a

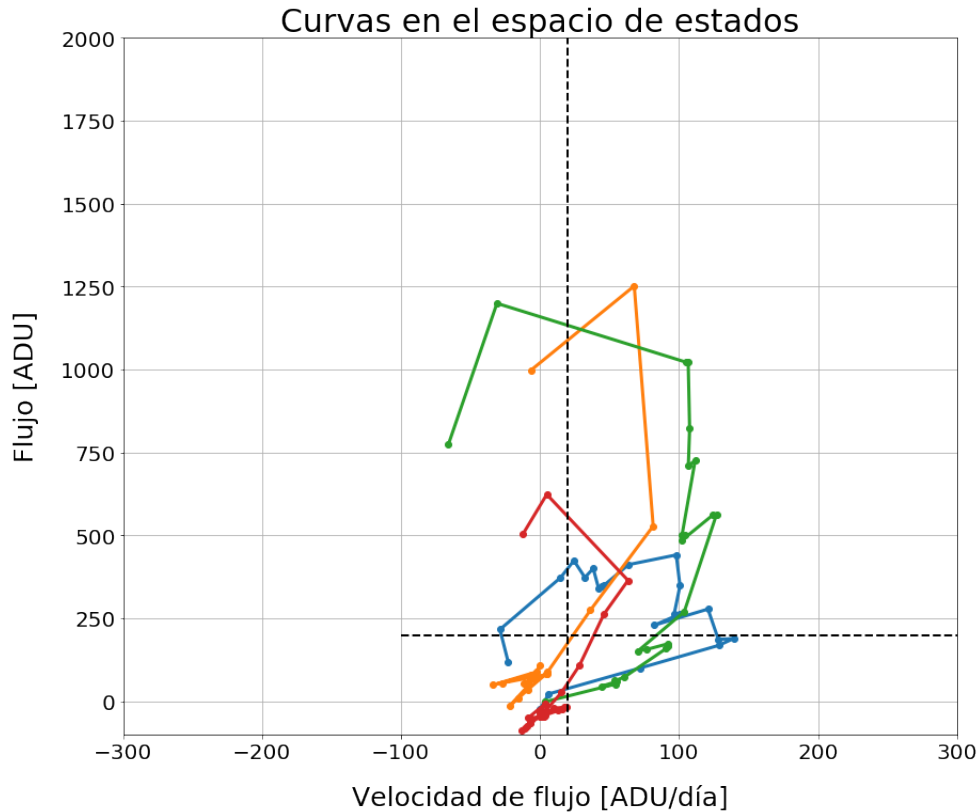


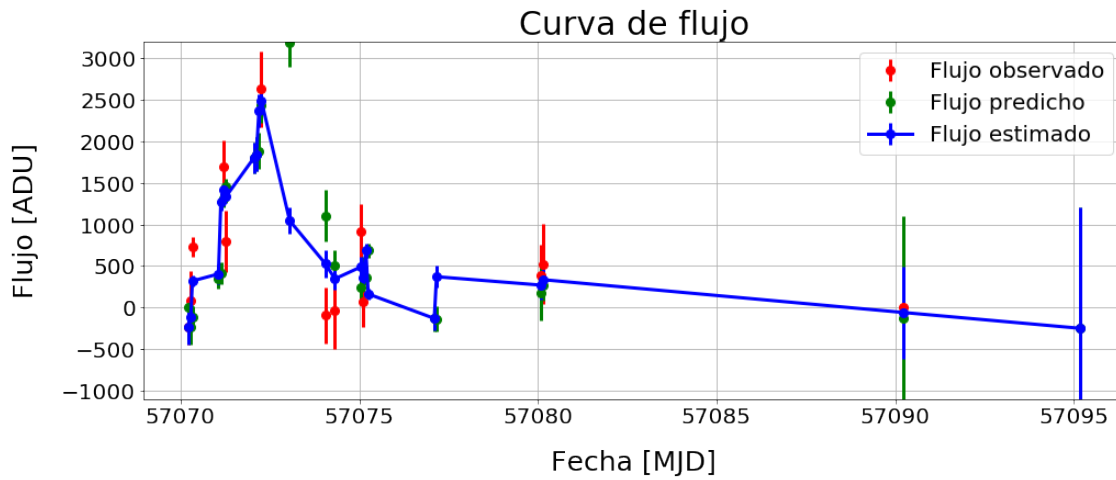
Figura 4.18: Ejemplos de curvas en sus espacio de estados de supernovas del conjunto de prueba, descubiertas solo por los modelos relajado, final y final podado. Además, se aprecian 2 líneas rectas discontinuas de color negro, las cuales corresponden a los umbrales de flujo (200 [ADU]) y de velocidad de flujo (20 [ADU/días]).

lo anterior, el programa emite una alerta de supernova.

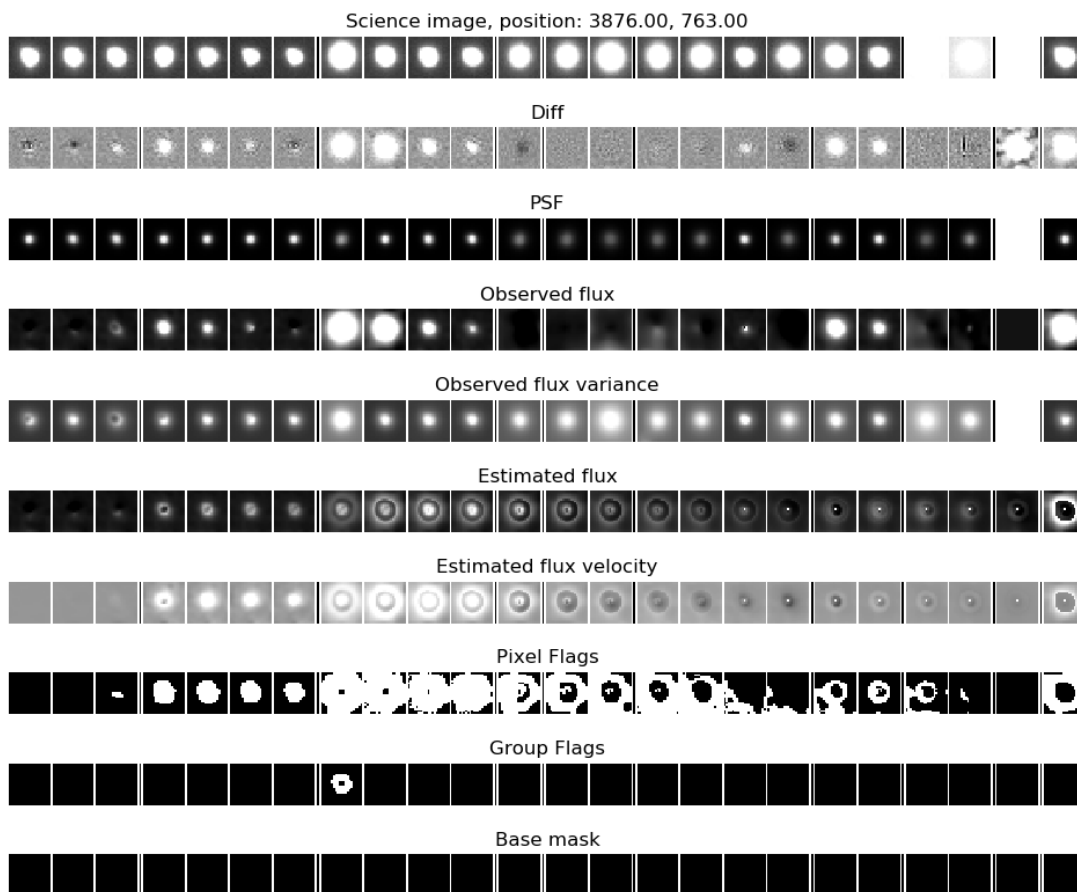
En la Figura 4.21 se observa uno de los falsos negativos. Esta supernova no fue descubierta dado que como se observa en su curva de flujo (ver Figura 4.21a), esta apareció al final de la campaña de HiTS 2015A. Esto es corroborado en la Figura 4.21b, donde se ve que en su imagen científica, de diferencia y de flujo observado solo aparecen elementos en las últimas 2 observaciones. La rutina propuesta requiere que la supernova aparezca al menos por 4 épocas consecutivas.

4.8.2. Imágenes filtradas según *airmass* y alta cadencia vs baja cadencia

En la Tabla 4.19 se observan 2 casos de interés, el primero evalúa el comportamiento del programa al utilizar todas las observaciones de las que se tenga información *versus* el comportamiento de utilizar solo las observaciones pertenecientes al periodo de mayor cadencia. Al comparar los modelos marcados como ID 1 e ID 2 se aprecia que al restringir el número de observaciones a solo la cantidad de alta cadencia, se aprecia una disminución en la cantidad de TP, FN y FP en los conjuntos de entrenamiento y validación, mientras que en el conjunto de prueba se mantiene la cantidad de TP, FN y FP. De igual forma, al comparar los modelos



(a) Curva de flujo de un falso positivo obtenido por el modelo final.



(b) Estampillas de un falso positivo obtenido por el modelo final.

Figura 4.19: Imágenes correspondiente a un falso positivo encontrado por el modelo final, el cual correspondería a un candidato a estrella variable.

identificados como ID 3 e ID 4 los resultados son análogos al caso anterior, solo que ahora disminuye la cantidad de TP en el conjunto de prueba. A partir de estos experimentos se muestra que al considerar solo las observaciones de alta cadencia no existe una disminución

considerable en los falsos positivos (en ningún conjunto ocurre un cambio importante) pero sí se ve perjudicado la cantidad de TP (sobre todo en los conjuntos de entrenamiento y validación de ambas comparaciones). Esto indicaría que el programa es capaz de manejar correctamente observaciones que no pertenezcan a un periodo de alta cadencia. Lo anterior se debe al σ_{modelo} con el que se trabajo, dado que este permite confiar más en la observación que en el modelo matemático propuesto (el cual tiende a fallar cuándo existen grandes diferencias de tiempo entre predicciones).

El otro caso de interés corresponde al estudio del *airmass*, en el cual se evalúa el comportamiento del programa al considerar series de imágenes filtradas de épocas cuyas imágenes científicas presentan un valor de *airmass* superior a 1,7 (definido por un experto). A través del par de experimentos de ID 1 e ID 3 es posible observar que a pesar de que se consideran todas las imágenes (sin umbral de *airmass*) los resultados parecen no variar demasiado. Se presenta un leve aumento (2 nuevas supernovas son detectadas) en el conjunto de prueba en el cual aumenta levemente también la cantidad de falsos positivos. Del mismo modo, a través del par de modelos asociados a los IDs 2 y 4 es posible observar un aumento en una unidad en la cantidad de TP del conjunto de validación y de prueba, manteniendo un número similar de falsos (levemente mayor) para el caso en el que se consideran todas las imágenes. Lo anterior se debe a que el filtro de Kalman de máxima correntropía es capaz de filtrar la información que presenta un *airmass* superior a 1,7 dado que son considerados como *outliers*.

Tabla 4.19: Comparación de resultados al utilizar el modelo final podado considerando 4 casos. En la columna **observaciones** se tienen dos opciones, el primero, **todas** que se refiere a que se consideran las observaciones de todo el periodo de observación, mientras que la segunda opción corresponde a **alta cadencia**, es decir, solo se considerarán las observaciones encontradas en el período de mayor cadencia. Por otro lado, la columna *airmass* se refiere a si se consideran o no las observaciones filtradas por este umbral.

			Entrenamiento			Validación			Prueba		
ID	Observaciones	<i>airmass</i>	TP	FN	FP	TP	FN	FP	TP	FN	FP
1	Todas	No	25	19	29	11	11	18	12	11	18
2	Alta cadencia	No	20	24	28	8	14	17	12	11	18
3	Todas	Si	25	19	32	11	11	37	14	9	23
4	Alta cadencia	Si	19	25	31	9	13	34	13	10	23

4.8.3. Comparación entre el uso de filtros de Kalman de máxima correntropía y básico

En la Tabla 4.20 se observan los resultados del modelo final al modificar su filtro. Los cambios en los 3 conjuntos son considerablemente bajos, la cantidad de TP en el conjunto de prueba del modelo con filtro de Kalman de máxima correntropía solo supera por una unidad a la cantidad de TP del conjunto de prueba del modelo con filtro de Kalman básico. La cantidad de falsos positivos logra disminuir en los 3 conjuntos, pero de forma poca significativa.

En la Figura 4.22 se observa la cantidad de veces que son detectadas las supernovas descubiertas del conjunto de prueba, donde claramente se aprecia que no existe diferencia entre el uso del modelo con filtro de Kalman de máxima correntropía y el modelo que utiliza

el filtro de Kalman básico. La única diferencia que se aprecia en el gráfico es que existe una supernova (SN41) que solo es detectada por el filtro MCC.

En la Tabla 4.21 se observa que para la mayoría de los casos (excepto supernova SN41) no existe ninguna diferencia en los tiempos de detección.

A través de las tablas y figura mostradas en esta sección, se aprecia que el modelo con filtro de Kalman básico posee un comportamiento bastante similar al modelo con filtro de Kalman de máxima correntropía. A partir de los resultados de estas pruebas, quedaría a criterio del experto que tipo de filtro utilizar. La gran ventaja de usar el filtro de Kalman básico es que es menos costoso computacionalmente que el filtro de Kalman de máxima correntropía, lo que reduciría los tiempos de ejecución del programa. Esto último, posee una gran importancia dado que si se quiere implementar un detector *online* se privilegiaría el que posea un buen comportamiento (que sea capaz de detectar una gran cantidad de supernovas) y que a su vez sea rápido de procesar. Por otro lado, a pesar que el filtro de Kalman de máxima correntropía resulta ser levemente mejor que el filtro básico (en cuanto a cantidad de TP y FP), se tendría que probar con una base de datos más grande para confirmar que respectivamente es mejor que el filtro de Kalman básico.

Tabla 4.20: Comparación de cantidad de TP, FN y FP obtenidos al utilizar el modelo final variando su filtro.

Modelo	Filtro de Kalman	Entrenamiento			Validación			Prueba		
		TP	FN	FP	TP	FN	FP	TP	FN	FP
Final	Básico	24	20	17	11	11	23	10	13	13
Final	Máxima Correntropía	24	20	13	11	11	16	11	12	10

Tabla 4.21: Diferencias de tiempo (en días) con respecto a HiTS de las supernovas detectadas por el modelo final con filtro de Kalman básico (Básico) *versus* el modelo final con filtro de Kalman de máxima correntropía (MCC). Si el valor es negativo significa que es detectado antes que la campaña HiTS, si es positivo significa que fue detectado después que la campaña HiTS, si es 0 significa que fue descubierto al mismo tiempo y es **x** si no fue descubierto por el modelo respectivo.

Filtro	Identificador de supernovas										
	44	26	25	13	28	27	32	41	14	35	51
Básico	-6,0	-2,0	0,1	0,0	2,0	-3,0	-4,1	x	2,0	0,0	-2,9
MCC	-6,0	-2,0	0,1	0,0	2,0	-3,0	-4,1	2,0	2,0	0,0	-2,9

4.8.4. Prueba con otra base de datos

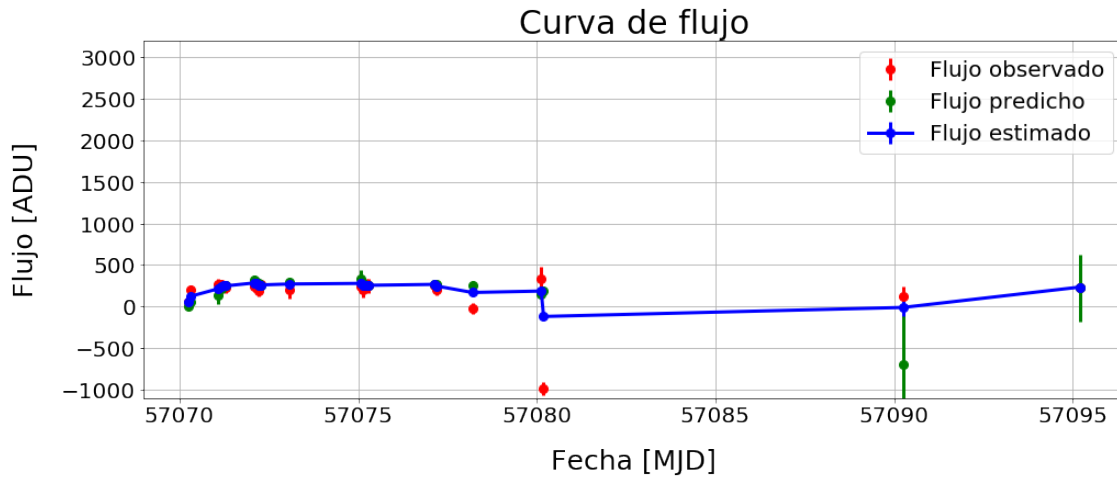
En la Tabla 4.22 se observa que al utilizar otra base de datos, el comportamiento del modelo final (no podado) y del modelo final podado, obtienen una mayor cantidad de detecciones que el modelo original. La cantidad de falsos positivos en el original es menor que para el resto de los casos, pero el modelo no podado de todas formas posee un número aceptable de estos.

A través de los resultados de la Tabla 4.22 se ilustra que los nuevos modelos generados son capaces de detectar una mayor cantidad de supernovas manteniendo en un número razonable

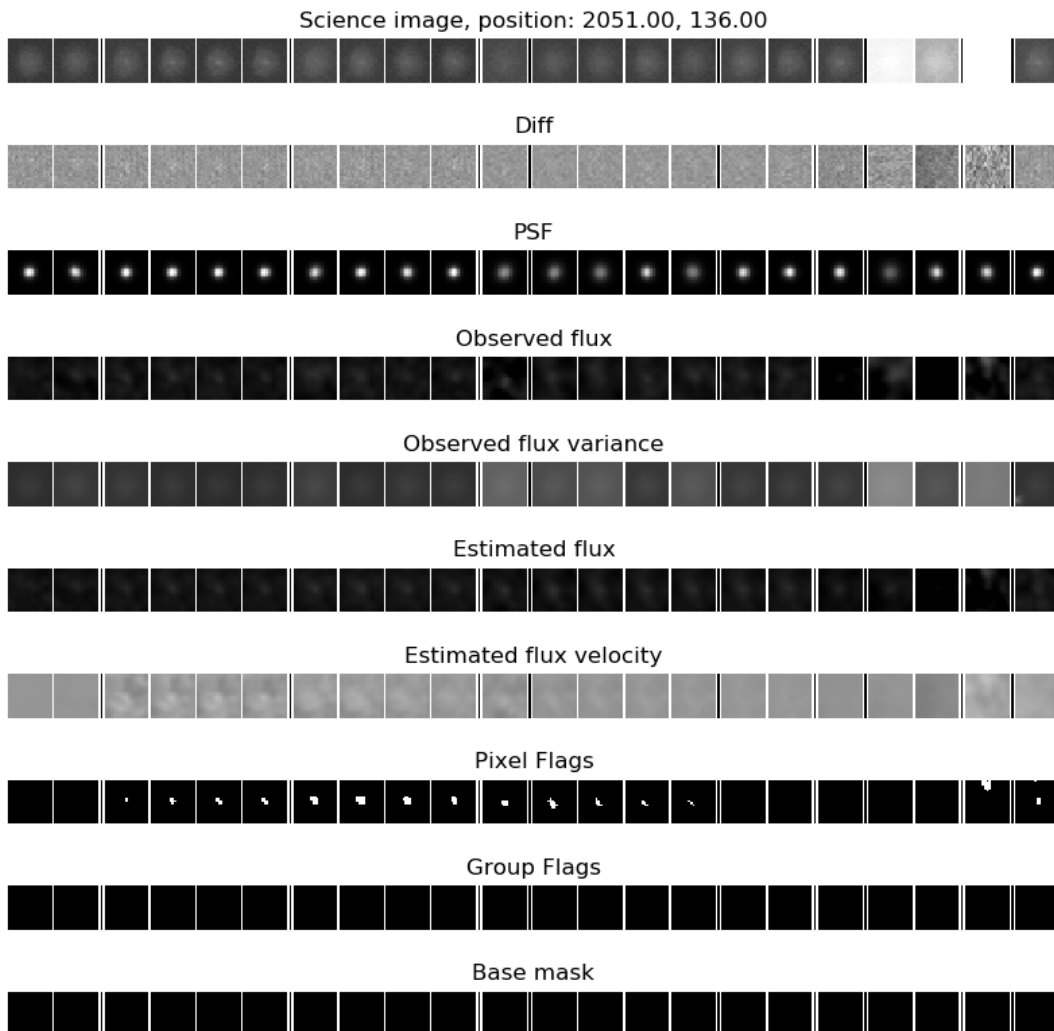
de falsos positivos. Dentro de estos, el modelo final podado presentó una gran cantidad de falsos positivos, a diferencia del modelo final (sin poda).

Tabla 4.22: Número de verdaderos positivos (TP), falsos negativos (FN) y falsos positivos (FP) detectados de la campaña del 2014A de HiTS utilizando los modelos original, final y final podado.

Modelo	Prueba		
	TP	FN	FP
Original	7	24	24
Final	17	14	38
Final podado	17	14	94

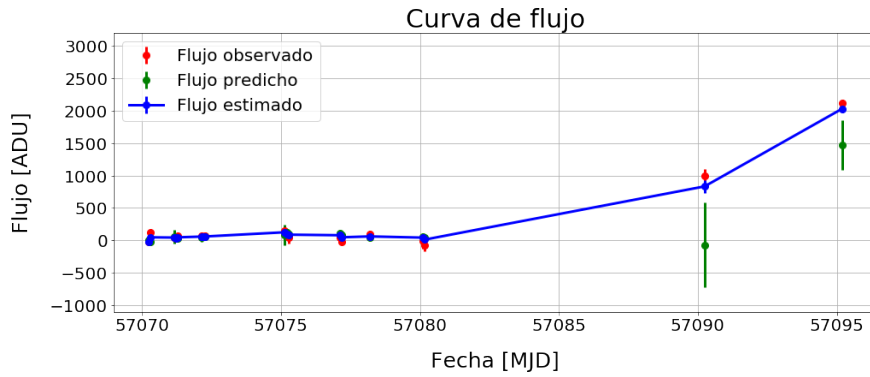


(a) Curva de flujo de un falso positivo obtenido por el modelo final.

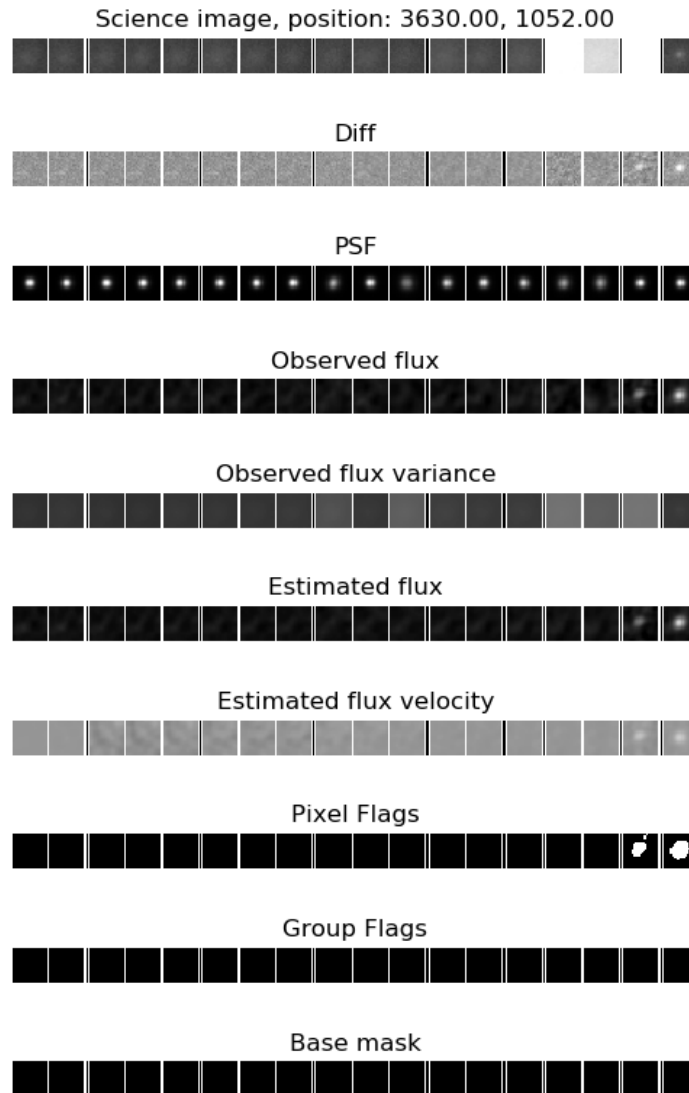


(b) Estampillas de un falso positivo obtenido por el modelo final.

Figura 4.20: Imágenes correspondiente a un falso positivo encontrado por el modelo final, el cual se produce por la existencia de un crecimiento aparente en su flujo.



(a) Curva de flujo de un falso negativo obtenido por el modelo final.



(b) Estampillas de un falso negativo obtenido por el modelo final.

Figura 4.21: Imágenes correspondiente a un falso negativo, el cual se produce por una observación tardía del objeto.

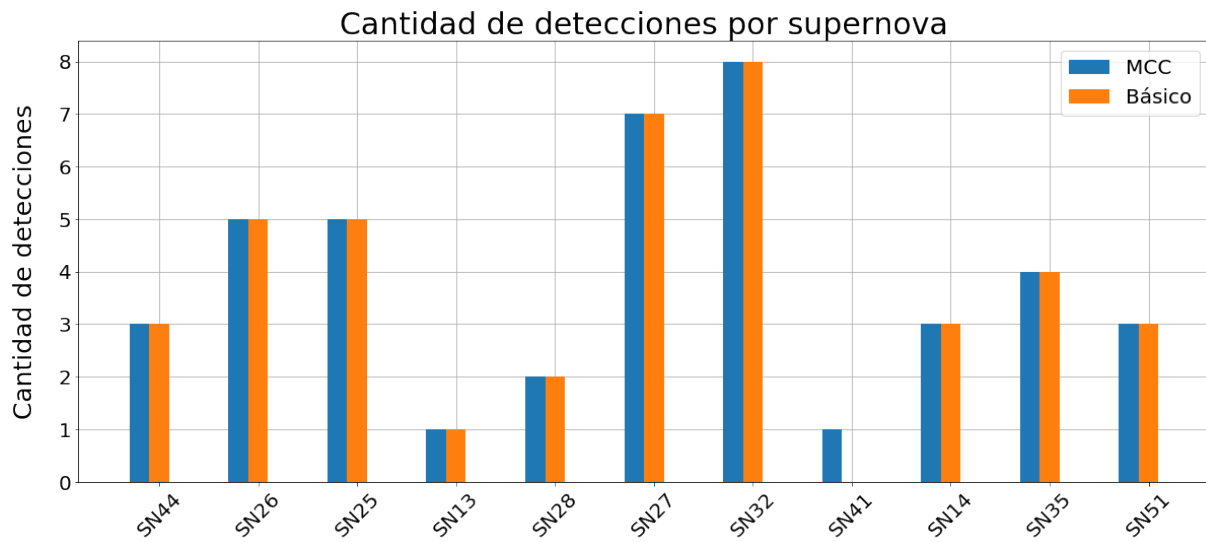


Figura 4.22: Cantidad de detecciones por supernova descubierta en el conjunto de prueba, comparando el modelo original con los modelos finales con distinto filtro.

Conclusión

En el presente trabajo se propuso mejorar la cantidad y calidad de detecciones de supernovas a partir del uso del filtro de Kalman de máxima correntropía y entropía de curvas.

Se desarrolló un método capaz de discriminar las alertas emitidas por la rutina original a través del cálculo de la entropía de la curva en el espacio de estados de éstas. Este método permitió trabajar sobre un modelo relajado, con el cual se detectó una mayor cantidad de TP que en el modelo original con una cantidad de FP similares. A través del establecimiento de un umbral de entropía, se logra reducir la cantidad de FP progresivamente a través de la disminución de dicho valor, logrando establecer que, para las pruebas realizadas en este trabajo, con un valor de 0,3 el modelo obtiene un buen *trade off* entre la cantidad de TP y FP. A pesar de esto, este parámetro es ajustable, es decir, en caso de que un experto acepte encontrar una mayor cantidad de falsos positivos a cambio de encontrar más supernovas, es posible aumentar dicho valor. Lo anterior mostró que la entropía de las curvas en el espacio de estados de las supernovas tienden a poseer valores más bajos que los objetos que no lo son.

Otra de las ventajas de trabajar sobre un modelo relajado, es que es posible detectar una mayor cantidad de veces una misma supernova, es decir, el programa es capaz de seguir el comportamiento de mejor forma que el modelo original. Además, el modelo relajado permitió descubrir las supernovas, épocas antes que el modelo original e incluso que HiTS. Estas ventajas no se vieron afectadas al implementar el método del cálculo de entropía de curvas.

Además, a través de la variación de los metaparámetros asociados al filtro de Kalman de máxima correntropía se logró disminuir (levemente) la cantidad de falsos positivos. Al aumentar el valor del parámetro σ_{modelo} , se eliminaron los falsos emitidos por malas estimaciones del filtro provocadas por las grandes diferencias de tiempo entre una observación y otra. Además, el parámetro $\sigma_{correntropia}$, fijado a través del método de prueba y error, no presentó cambios considerables entre los valores que se probaron a excepción de cuando éste era muy bajo ($\sigma_{correntropia} = 10$) el cual disminuyó la cantidad de detecciones correctas y aumentó la cantidad de falsos.

A través de la realización de una poda de las reglas de la rutina original, se determinó que el programa, con sus nuevas modificaciones (como por ejemplo, implementación del cálculo de entropía de la curva de espacio de estados), es capaz de mantener buenos resultados y comparables con el modelo sin poda. Esto ayudará a futuros trabajos dado que al poseer menos reglas, se tienen menos parámetros que analizar, por lo que es posible realizar un análisis más exhaustivo del resto. Además, esto permitiría evitar problemas de sobreajuste,

los cuales ocurren cuando un programa considera una gran cantidad de reglas y/o condiciones.

El enfoque de este trabajo fue determinar un modelo capaz de detectar una mayor cantidad de supernovas que las obtenidas por el modelo original, manteniendo una cantidad similar o inferior de falsos positivos. Si bien es cierto, este enfoque permitiría al astrónomo realizar una menor cantidad de revisiones visuales, aún existe una gran cantidad de falsos negativos, es decir, supernovas que no logran ser detectadas. Una forma de detectar un mayor número de supernovas (a costo de encontrar más falsos positivos) sería aumentando el valor del umbral de entropía, el cual puede ser ajustado a criterio del experto.

A pesar que el aumento en la cantidad de supernovas detectadas del modelo final, en comparación al modelo original, es relativamente bajo, se debe considerar que la base de datos era pequeña, por lo que se esperaría que al aumentar la cantidad de datos esta diferencia sería más considerable. Además se debe recordar que el modelo sería capaz de encontrar más supernovas al ajustar el umbral de entropía (a costo de obtener una mayor cantidad de falsos).

El aumento en la cantidad de detecciones no fue la única ventaja que se obtuvo al trabajar el detector de supernovas. Además, el modelo final fue capaz de detectar supernovas antes que HiTS, lo cual permitiría estudiarlas desde más temprano y así obtener datos de interés astronómico. Además, logra realizar un seguimiento de estos objetos de forma más consistente.

4.9. Trabajo futuro

En este trabajo se mostró que la implementación de la entropía de curvas fue capaz de discriminar una gran cantidad de falsos positivos perdiendo, en el peor de los casos, un par de verdaderos positivos. Se propone implementar esta condición dentro del mismo programa y no dejarla como un filtro externo como se hizo actualmente. Además de estudiar la secuencia de valores de entropía y no estudiarlos individualmente.

Uno de los problemas de este trabajo es que el modelo matemático posee supuestos demasiado fuertes (como que la velocidad es constante entre dos mediciones). Es por esto que se propone estudiar nuevas ecuaciones que describan el comportamiento de este parámetro para recopilar información que se acerque al comportamiento real.

Se proponen también otros métodos de discriminación de supernovas, como lo es el estudio de la forma de las curvas en el espacio de estados. Se espera comparar punto a punto una supernova completamente nueva con las ya confirmadas y conocidas, donde si los puntos poseen un comportamiento similar a alguna supernova ya conocida, ésta sea propuesta como supernova.

A pesar de que se realizó un estudio del sigma del ancho de banda del kernel Gaussiano para el cálculo de la entropía ($\sigma_{correntropia}$), éste no presentó cambios considerables en la cantidad de objetos discriminados. Es por esto que se propone crear un $\sigma_{correntropia}$ adaptativo, es decir, variar este parámetro dependiendo de las imágenes de entrada. Además, para establecer este parámetro se propone el uso del método de Silverman, el cual permite encontrar un $\sigma_{correntropia}$ candidato para realizar un correcto ajuste.

Otro posible trabajo futuro sería el de retomar el estudio del filtro de Kalman *Unscented*, el cual fue iniciado en el trabajo de Paloma Pérez [7] y en este trabajo de memoria no fue abordado.

Dado que se espera utilizar este método en otras bases de datos, se propone adecuar el código para que funcione con los datos que éstas entreguen. Una forma de realizar esto sería trabajando y modificando las reglas propuestas por la rutina original, donde se deberían eliminar y/o modificar para que éstas sean coherentes con la información entregada. Además de esto, se debe adecuar la rutina al tamaño de imágenes e información extra que se obtenga.

Bibliografía

- [1] N. Kaiser, H. Aussel, B. E. Burke, H. Boesgaard, K. Chambers, M. R. Chun, J. N. Heasley, K.-W. Hodapp, B. Hunt, R. Jedicke, *et al.*, “Pan-starrs: a large Synoptic Survey Telescope Array”, in *Proceedings of SPIE - The International Society for Optical Engineering*, vol. 4836, pp 154-165, January 2002.
- [2] W. Sutherland, J. Emerson, G. Dalton, E. Atad-Ettedgui, S. Beard, R. Bennett, *et al.*, “The Visible and Infrared Survey Telescope for Astronomy (VISTA): Design, Technical Overview and Performance”, in *Astronomy & Astrophysics*, vol. 575, March 2015.
- [3] P. A. Abell, J. Allison, and S.F. Anderson. LSST Science Collaborations *et al.* arXiv preprint arXiv:0912.0201, 2009.
- [4] T. Diehl, “The Dark Energy Survey Camera (DECam)”, in *Physics Procedia*, vol. 37, pp. 1332–1340, January 2012.
- [5] F. Förster, J. C. Maureira, J. San Martín, M. Hamuy, J. Martínez *et al.*, “The High Cadence Transient Survey (HITS). I. Survey Design and Supernova Shock Breakout Constraints,” in *The Astrophysical Journal*, vol. 832, p. 155, December 2016.
- [6] P. Huentelmu, “Filtro de Correntropía para detección de supernovas”, Tesis de magíster en ciencias de la ingeniería, mención eléctrica, Departamento de Ingeniería Eléctrica en Facultad de Ciencias Físicas y Matemáticas, Universidad de Chile, Santiago, 2016.
- [7] P. Pérez, “Extensión de filtro de Kalman de aproximación no lineal para la detección de objetos astronómicos”, Memoria para optar al título de ingeniería civil en computación, Departamento Ciencias de la Computación en Facultad de Ciencias Físicas y Matemáticas, Universidad de Chile, Santiago, 2019.
- [8] R. E. Kalman, “A new approach to linear filtering and prediction problems”, in *Journal of Basic Engineering*, vol. 82, pp. 35-45, March 1960.
- [9] A. Gunduz, A.Hedge, J.C. Príncipe, “Correntropy as a Novel Measure for Nonlinearity Tests”, in *The 2006 IEEE International Joint Conference on Neural Network Proceedings*, Vancouver, Canada, 2006.
- [10] B. Chen, X. Liu, H. Zhao, and J. C. Principe, “Maximum correntropy kalman filter”, in *Automatica*, vol. 76, pp. 70-77, 2017.

- [11] A. Balestrino, A. Caiti, and E. Crisostomi, “Generalised entropy of curves for the analysis and classification of dynamical systems”, in *Entropy*, vol. 11(2). pp. 249-270, 2009.
- [12] T. Naylor, “An optimal extraction algorithm for imaging photometry”, in *Monthly Notices of the Royal Astronomical Society*, vol. 296, no. 2, pp. 339–346, May 1998.
- [13] Jose C. Principe, *Information Theoretic Learning: Renyi’s Entropy and Kernel Perspectives*, Springer Publishing Company, Incorporated, 2010.
- [14] R. Carrasco, “Image sequence simulation and deep learning for astronomical object classification”, Tesis de magíster en ciencias de la ingeniería, mención eléctrica, Departamento de Ingeniería Eléctrica de la Facultad de Ciencias Físicas y Matemáticas, Universidad de Chile, Santiago, 2019.

Apéndice A

Reglas de rutina original

A.1. Reglas individuales

Reglas que deben superar individualmente los píxeles de cada imagen para poder ser considerado como un píxel candidato.

1. Valor del flujo estimado por el Filtro de Kalman del píxel debe ser mayor a un umbral dado.
2. Valor de la velocidad de flujo estimado por el Filtro de Kalman del píxel debe ser mayor a un umbral dado.
3. Valor de la imagen científica del píxel es menor a la mediana de dicha imagen más un delta.
4. Varianza del flujo del píxel es mayor a un umbral dado.
5. Varianza de la velocidad del flujo del píxel es mayor a un umbral dado.
6. Si los píxeles no son considerados defectuosos (máscara).
7. Si los píxeles no han sido descartados durante n (valor original corresponde a $n = 4$) épocas consecutivas por superar la mediana estimada.

A.2. Reglas grupales

Una vez determinado los píxeles individuales que podrían formar parte de una supernova se proceden a agrupar para posteriormente obtener información de su centro. Estas reglas se enfocaran en la información central de cada grupo y su alrededor.

1. Se descarta grupo si posee una mala resta alrededor del centro del píxel en estudio dentro de un sector.
2. Si no existen máximos locales en la imagen científica dentro del grupo de píxeles en estudio.
3. Si no existen máximos locales en la imagen de flujo, dentro del grupo en estudio, obtenida al aplicar fotometría de Naylor.

4. Si no existen máximos locales dentro de la imagen de velocidad de flujo, dentro del grupo en estudio, obtenida al aplicar fotometría de Naylor.
5. Si los valores del grupo superan la mediana local de la imagen científica.
6. Si el grupo de píxeles en estudio posee algún píxel que doble el valor de la imagen científica o de la de flujo.
7. Si el centro del grupo se encuentra dentro de la máscara de píxeles defectuosos.
8. Si el píxel central del grupo no ha sido descartado durante n (valor original corresponde a $n = 4$) épocas consecutivas por superar la mediana estimada.
9. Si la varianza de flujo del píxel central es mayor a un umbral dado.

Apéndice B

Resultados y análisis

B.1. Tablas

Tabla B.1: Número de verdaderos positivos (TP), falsos negativos (FN) y falsos positivos (FP) detectados por el modelo original con el factor γ (Con *Zero Points*) y sin éste (Sin *Zero Points*).

Modelo	Entrenamiento			Validación			Prueba		
	TP	FN	FP	TP	FN	FP	TP	FN	FP
Sin Zero Points	19	25	11	9	13	18	8	15	8
Con Zero Points	17	27	19	9	13	25	8	15	12

Tabla B.2: Número de verdaderos positivos (TP), falsos negativos (FN) y falsos positivos (FP) detectados, obtenidos al variar los parámetros *accum neg flux depth* (**neg**) y *accum med flux depth* (**med**).

Parámetros		Entrenamiento			Validación		
neg	med	TP	FN	FP	TP	FN	FP
4	3	19	25	11	9	13	19
1	1	27	17	147	11	11	78

Tabla B.3: Número de verdaderos positivos (TP), falsos negativos (FN) y falsos positivos (FP) detectados, obtenidos al variar el parámetro de varianza de flujo.

Varianza de flujo	Entrenamiento			Validación		
	TP	FN	FP	TP	FN	FP
250	30	14	321	14	8	167
500	30	14	341	14	8	176

Tabla B.4: Número de verdaderos positivos (TP), falsos negativos (FN) y falsos positivos (FP) detectados, variando número de puntos para el cálculo del *convex hull*.

	Entrenamiento			Validación		
Lim	TP	FN	FP	TP	FN	FP
6	28	16	168	13	9	116
10	29	15	102	12	10	85
14	29	15	76	12	10	71
16	29	15	70	12	10	62
18	29	15	70	12	10	62

Tabla B.5: Número de verdaderos positivos (TP), falsos negativos (FN) y falsos positivos (FP) detectados, obtenidos al variar el umbral de entropía. Se utilizó como base el modelo relajado.

	Entrenamiento			Validación		
Umbral de Entropía	TP	FN	FP	TP	FN	FP
0,3	24	20	13	11	11	18
0,6	26	18	44	11	11	53

Tabla B.6: Número de verdaderos positivos (TP), falsos negativos (FN) y falsos positivos (FP) detectados, variando $\sigma_{correntropia}$ usando como base el modelo relajado.

	Entrenamiento			Validación		
$\sigma_{correntropia}$	TP	FN	FP	TP	FN	FP
1000	30	14	341	14	8	176
400	30	14	346	14	8	177
200	30	14	310	14	8	150
10	19	25	361	8	14	187

Tabla B.7: Número de verdaderos positivos (TP), falsos negativos (FN) y falsos positivos (FP) detectados, realizando poda de condiciones.

	Entrenamiento			Validación		
N° de condiciones eliminadas	TP	FN	FP	TP	FN	FP
0	24	20	13	11	11	16
1	24	20	13	11	11	16
2	24	20	13	11	11	16
3	24	20	13	11	11	16
4	24	20	15	11	11	17
5	24	20	21	11	11	17
6	24	20	22	11	11	18
7	25	19	29	11	11	18