



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO  
POSGRADO EN ASTROFÍSICA  
INSTITUTO DE ASTRONOMÍA

**METODOLOGÍA PARA LA DETECCIÓN AUTOMÁTICA  
DE CANDIDATOS A TRÁNSITOS DE EXOPLANETAS  
EN CURVAS DE LUZ DE ALTA CADENCIA**

T E S I S  
QUE PARA OBTENER EL TÍTULO DE  
MAESTRO EN CIENCIAS (ASTROFÍSICA)

PRESENTA:  
FERNANDO IVAN ALVAREZ SANTANA

DIRECTORES:  
DR. JOEL H. CASTRO CHACÓN                      DR. MAURICIO REYES RUIZ  
INSTITUTO DE ASTRONOMÍA, UNAM.

Ensenada, Baja California, México. a 13 de agosto del 2020



Universidad Nacional  
Autónoma de México



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

## Agradecimientos

Quiero agradecer a todo el equipo de TAOS II en el Instituto de Astronomía de la UNAM sede Ensenada, por sus comentarios y enseñanzas. En especial quiero agradecer a mis directores de tesis Joel Humberto Castro y Mauricio Reyes Ruiz sin los cuales, este trabajo no hubiera sido posible.

Agradezco también al Consejo Nacional de Ciencia y Tecnología (CONACYT) y a los proyectos PAPIIT IN110217 y IN107316 por el apoyo económico para realizar mis estudios de maestría y diversas actividades.

## Índice general

<b>Lista de figuras</b> . . . . .	X
<b>Resumen</b> . . . . .	XI
<b>I. Introducción</b> . . . . .	1
I.1 Contexto histórico . . . . .	1
I.2 Proyectos enfocados en la búsqueda de exoplanetas por el método de tránsito . . . . .	3
I.2.1 CoRoT . . . . .	3
I.2.2 HATNet y HATSouth . . . . .	5
I.2.3 SuperWASP . . . . .	5
I.2.4 Telescopio espacial <i>Kepler</i> . . . . .	6
I.2.5 Satélite <i>TESS</i> . . . . .	8
I.3 Proyecto TAOS II . . . . .	9
I.4 Objetivos, Hipótesis y estructura de la tesis . . . . .	11
<b>II. Marco teórico</b> . . . . .	12
II.1 Señales observadas por tránsitos de exoplanetas . . . . .	12
II.1.1 Sistema general de ecuaciones . . . . .	12
II.1.2 Efecto del oscurecimiento del limbo en las curvas de luz . . . . .	15
II.1.3 Falsos positivos en la búsqueda de tránsitos . . . . .	16
II.2 Observación convencional de tránsitos de exoplanetas . . . . .	18
II.3 Reducción de imágenes en curvas de luz . . . . .	19
II.3.1 Preprocesamiento de imágenes . . . . .	21
II.3.2 Fotometría rápida . . . . .	22
II.3.1 Fotometría de apertura . . . . .	23
II.3.2 Fotometría diferencial . . . . .	25

---

II.4	Análisis de ruido en series de tiempo . . . . .	26
II.5	Transformada de Fourier para estudiar tiempo y frecuencia . . . . .	28
<b>III.</b>	<b>Metodología . . . . .</b>	<b>31</b>
III.1	Datos observacionales, estudios preparatorios para TAOS II . . . . .	31
III.1.1	Telescopio e instrumentos en la temporada de observación . . . . .	32
III.1.2	Descripción de objetos observados . . . . .	33
III.1.3	Descripción de datos, influencia del seeing y ruido sistemático . . . . .	33
III.2	Análisis de datos observacionales . . . . .	35
III.2.1	Fotometría con APPHi . . . . .	35
III.3	Análisis de ruido en curvas de luz de alta cadencia . . . . .	37
III.3.1	Distribución temporal del ruido . . . . .	37
III.3.2	Espectro de frecuencias del ruido . . . . .	41
III.3.3	Simulación de ruido . . . . .	43
III.4	Simulación de tránsitos de exoplanetas . . . . .	48
III.4.1	Descripción y uso de BATMAN . . . . .	48
III.5	Reducción de ruido en curvas de luz . . . . .	49
III.5.1	Reducción del ruido usando promedio móvil . . . . .	50
III.5.2	Identificación de firma de tránsito usando filtro pasa bajas . . . . .	52
III.6	Criterio de validación para candidatos a tránsitos de exoplanetas . . . . .	56
III.6.1	Modelo del trapezoide . . . . .	56
III.6.2	Coefficiente de correlación de Pearson . . . . .	57
<b>IV.</b>	<b>Análisis de Resultados . . . . .</b>	<b>58</b>
IV.1	Simulaciones . . . . .	58
IV.1.1	Determinación de candidatos a tránsitos . . . . .	58
IV.2	Datos observacionales . . . . .	66
IV.2.1	Mejora de la señal a ruido en datos observacionales . . . . .	66
IV.2.2	Determinación de candidatos a tránsitos . . . . .	73
<b>V.</b>	<b>Conclusiones . . . . .</b>	<b>77</b>

<b>Anexos</b> . . . . .	79
6.1. Datos observacionales . . . . .	79
6.2. Paquetería HCLcat . . . . .	79
Paquetería HCLcat . . . . .	79

## Índice de figuras

2.1.	Esquema de un tránsito de exoplaneta observable. $i$ representa la inclinación de la órbita y $R_*$ es el radio de la estrella. $b$ es el parámetro de impacto, esta cantidad adimensional es utilizada para resumir la información de la inclinación de la órbita y el semi eje de la órbita $a$ . . . . .	12
2.2.	Proceso que ocurre durante un tránsito planetario. La curva en la parte inferior representa el flujo de la estrella medido por el observador, $i$ representa la inclinación de la órbita, $R_*$ es el radio de la estrella, $R_p$ el radio del planeta, $\Delta F$ es el cambio en el flujo visto por el observador, $t_T$ es la duración total del tránsito y $b$ el parámetro de impacto. . . . .	15
2.3.	Efecto del oscurecimiento del limbo (LD) en curvas de luz de tránsitos simulados. Se muestran 4 modelos distintos, en azul tenemos LD constante, en naranja un modelo lineal, en verde un modelo cuadrático y en rojo un modelo no-lineal. . . . .	16
2.4.	Visualización de los distintos casos más comunes de falsos positivos. Imagen tomada de la presentación de Raphaëlle D. Haywood, durante el 2018 Sagan Exoplanet Workshop (liga). . . . .	18
2.5.	Curva de luz de XO-3, tomada con el filtro R de Johnson en el telescopio de 84 cm en el Observatorio Astronómico Nacional (OAN-SPM) el 21 de octubre del 2016. . . . .	19
2.6.	Ejemplo de la imagen de una estrella estudiada en este trabajo, obtenida con una cámara rápida IXON en el telescopio de 84 cm en el OAN-SPM. . . . .	20
2.7.	Ejemplo de aperturas circulares para realizar fotometría de HAT-P-37b. El anillo azul, formado por las circunferencias $D_t$ y $A_t$ , se utiliza para calcular la contribución del flujo debido al cielo $S_t$ . . . . .	24

---

2.8.	Curva de luz de alta cadencia de Kepler-17b obtenida mediante fotometría de apertura. La observación de Kepler-17b se realizó en el sitio 2 de TAOS II, utilizando una cámara <i>Zyla</i> de Andor. Esta curva de luz se obtuvo utilizando APPHi véase (sección III.4.1). . . . .	25
2.9.	Curva de luz de alta cadencia de Kepler-17b obtenida mediante fotometría diferencial. La observación de Kepler-17b se realizó en el sitio 2 de TAOS II, utilizando una cámara <i>Zyla</i> de Andor. Esta curva de luz se obtuvo utilizando APPHi véase (sección III.4.1). . . . .	26
2.10.	Arriba: Curva de luz de alta cadencia de HAT-p-14b, tomada a 20Hz, con el telescopio de 84cm del OAN. Abajo: mostramos su espectro de frecuencias (FFT), con la amplitud en escala logarítmica para mejorar la visualización.	29
3.1.	Telescopio de 84 cm, en el OAN-SPM, cámara rápida iXon y rueda de filtros MEXMAN. . . . .	32
3.2.	SNR en función de la magnitud en el filtro R. Los puntos representan la SNR medida utilizando curvas de luz obtenidas en TAOS I. Las líneas sólidas representan la predicción de SNR utilizando ruido con distribución de Poisson, ruido del cielo y ruido de lectura. Gráfica tomada de (Lehner et al., 2016). . . . .	34
3.3.	Diagrama de flujo para ambos programas ( <i>APPHi field</i> y <i>APPHi star</i> ), donde D, B, F y Z representan los darks, bias, flats e imágenes de estudio respectivamente. Imágen tomada de (Sánchez et al., 2019). . . . .	36
3.4.	Arriba: Ejemplo de ruido blanco (gaussiano). Abajo: Histograma del ruido generado. . . . .	38
3.5.	Curva de luz de WASP-74b, calculada mediante fotometría diferencial usando APPHi. . . . .	39
3.6.	Arriba: Los puntos rojos son la distribución temporal de la curva de luz de WASP-74b. La línea azul sólida, es $m(t)$ resultado de un ajuste no lineal utilizando el algoritmo de Levenberg-Marquardt. Abajo: se aprecia los residuales de la diferencia entre la distribución y el modelo. . . . .	40
3.7.	Se muestra la contribución de las componentes $G_1$ y $G_2$ y la suma $m(t)$ . . .	41



---

3.8.	Arriba: se muestra una curva de ruido blanco con dos componentes $G_1$ y $G_2$ . Abajo: el espectro de frecuencias de la curva de luz, el cual permanece plano sin importar el número de componentes gaussianas. . . . .	42
3.9.	Arriba: se muestra una curva de luz de WASP-74b. Abajo: la zona de bajas frecuencias del espectro de la curva de luz, la línea negra es el ajuste de la envolvente exponencial $E(f)$ del espectro. La línea roja indica el espectro del tránsito inmerso en la curva de luz. . . . .	44
3.10.	Arriba: se muestra la FFT de la curva de luz de WASP-74b. Abajo: el espectro de frecuencias de la serie de tiempo $R(t)$ generada utilizando la metodología planteada en esta sección. Ambas gráficas son un acercamiento a la zona de bajas frecuencias para mejorar la visualización. La línea roja es la misma envolvente exponencial $E(f)$ para ambos espectros. . . . .	47
3.11.	Curvas de luz de tránsitos de exoplanetas con $P = 1$ día, $i = 90^\circ$ , $a = 15R_*$ , $\epsilon = 0$ , $\bar{\omega} = 90$ y un modelo de oscurecimiento lineal. Las distintas curvas representan la diferencia en el cambio del flujo $\Delta F$ para diferentes valores de $R_p$ . . . . .	49
3.12.	Curva de luz de WASP-74 tomada el día 28 de mayo del 2017. En azul se muestra la curva de luz original obtenida mediante fotometría de apertura. En rojo la curva resultado del promedio móvil, y en negro la curva resultado de tomar la mediana en lugar del promedio. En el recuadro se muestra la SNR calculada para cada curva de utilizando (2.11). . . . .	51
3.13.	Filtro pasa-bajas $\mathcal{H}(f)$ utilizado para remover altas frecuencias. $FC$ representa la frecuencia de corte la cual puede variar dependiendo la curva de luz. . . . .	53
3.14.	Curva de luz de WASP-74 tomada el día 28 de mayo del 2017. En azul se muestra la curva de luz original obtenida mediante fotometría de apertura. En rojo la curva resultado del filtrado de Fourier, utilizando una frecuencia de corte $FC = 2 \times 10^{-3}$ . . . . .	54
4.1.	Matriz de correlaciones de los parámetros $\Delta F$ y $t_c$ entre las simulaciones y los calculados después de aplicar el modelo del trapecoide a los datos filtrados utilizando Fourier. Estos resultados se obtuvieron a partir de curvas simuladas con una $SNR = 10$ y $SNR = 100$ . . . . .	59

- 
- 4.2. Las líneas sólidas representan la diferencia porcentual entre  $t_c^{Fit}$  y  $t_c^{Real}$  después de aplicar el modelo del trapecoide a los datos filtrados utilizando el filtro pasa bajas. Las líneas punteadas representan la misma cantidad, pero obtenida aplicando el modelo del trapecoide a las curvas simuladas sin filtrar. El color representa diferentes valores de  $\Delta F$ , se utilizaron 4%, 2% y 0.2% representados en amarillo, verde y rojo respectivamente. Estos resultados se obtuvieron a partir de curvas simuladas a las que se les insertó un ruido con una  $SNR = 10$  (izquierda) y  $SNR = 100$  (derecha). . . . . 61
- 4.3. Matriz de correlaciones de los parámetros  $\Delta F$  y  $t_c$  entre las simulaciones y los calculados después de aplicar el modelo del trapecoide a los datos filtrados utilizando promedio móvil. Estos resultados se obtuvieron a partir de curvas con una  $SNR = 10$  y  $SNR = 100$ . . . . . 62
- 4.4. Las líneas sólidas representan la diferencia porcentual entre  $t_c^{Fit}$  y  $t_c^{Real}$  después de aplicar el modelo del trapecoide a los datos filtrados utilizando promedio móvil. Las líneas punteadas representan la misma cantidad, pero obtenida aplicando el modelo del trapecoide a las curvas simuladas sin filtrar. El color representa diferentes valores de  $\Delta F$ , se utilizaron 4%, 2% y 0.2% representados en amarillo, verde y rojo respectivamente. Estos resultados se obtuvieron a partir de curvas simuladas a las que se les insertó un ruido con una  $SNR = 10$ . . . . . 63
- 4.5. Matriz de correlaciones de los parámetros  $\Delta F$  y  $t_c$  entre las simulaciones y los calculados después de aplicar el modelo del trapecoide a los datos filtrados utilizando PCA. Estos resultados se obtuvieron a partir de curvas con una  $SNR = 10$  y  $SNR = 100$ . . . . . 64
- 4.6. Las líneas sólidas representan la diferencia porcentual entre  $t_c^{Fit}$  y  $t_c^{Real}$  después de aplicar el modelo del trapecoide a los datos filtrados utilizando PCA. Las líneas punteadas representan la misma cantidad, pero obtenida aplicando el modelo del trapecoide a las curvas simuladas sin filtrar. El color representa diferentes valores de  $\Delta F$ , se utilizaron 4%, 2% y 0.2% representados en amarillo, verde y rojo respectivamente. Estos resultados se obtuvieron a partir de curvas simuladas a las que se les insertó un ruido con una  $SNR = 10$  (izquierda) y  $SNR = 100$  (derecha). . . . . 65

4.7.	Curva de luz de HAT-P-37b tomada a 20 fps, durante la segunda mitad del tránsito. La curva de luz se obtuvo mediante fotometría de apertura con una $SNR = 11.73$ . La línea negra representa el modelo teórico, utilizando los parámetros presentados en (Bakos et al., 2012). . . . .	67
4.8.	Azul: curva de luz de HAT-P-37b después del filtrado de altas frecuencias. La señal a ruido resultante fue de $SNR = 124.75$ . La línea negra representa el modelo teórico, utilizando los parámetros presentados en (Bakos et al., 2012). El coeficiente de correlación entre la curva y el modelo es $C = 0.982$ . La región sombreada representa el egreso del planeta del disco solar, la duración de este evento es $t_c$ . . . . .	68
4.9.	Azul: curva de luz de HAT-P-37b después del promedio móvil. La señal a ruido resultante fue de $SNR = 116.64$ . La línea negra representa el modelo teórico, utilizando los parámetros presentados en (Bakos et al., 2012). El coeficiente de correlación entre la curva y el modelo es $C = 0.962$ . La región sombreada representa el egreso del planeta del disco solar, la duración de este evento es $t_c$ . . . . .	69
4.10.	Azul: curva de luz de HAT-P-37b después del filtrado con PCA. La señal a ruido resultante fue de $SNR = 135.51$ . La línea negra representa el modelo teórico, utilizando los parámetros presentados en (Bakos et al., 2012). El coeficiente de correlación entre la curva y el modelo es $C = 0.889$ . La región sombreada representa el egreso del planeta del disco solar, la duración de este evento es $t_c$ . . . . .	70
4.11.	Arriba: Ruido simulado con componentes de baja frecuencia en su espectro. Abajo: en azul se muestra el resultado de PCA a una curva sin tránsito inmerso. La línea negra representa el modelo teórico de HAT-P-37b, utilizando los parámetros presentados en (Bakos et al., 2012). . . . .	72
4.12.	Arriba: Espectro de la curva de luz de WASP-74b del 06 de agosto. Abajo: Espectro de la curva de luz de HAT-P-37b del 06 de agosto. La línea roja es el espectro del tránsito simulado utilizando los parámetros presentados en (Bakos et al., 2012). La línea negra representa la envolvente exponencial del espectro. . . . .	75

---

6.1. Se muestran todas las 13 curvas de luz, a partir de las cuales se encontraron los resultados presentados en la sección IV.2. Las curvas en rojo, son las curvas de luz para las cuales no pudimos detectar el tránsito inmerso en el ruido. Las curvas azules son las observaciones donde si se detectó un tránsito, esto de acuerdo a la tabla 4.1. . . . . . 80

## Resumen

La detección y caracterización de exoplanetas ha tenido un gran crecimiento en los últimos años, proyectos como *Kepler* nos brindaron una visión amplia de las posibilidades cuando de encontrar sistemas planetarios se refiere. La sonda espacial *TESS* mapeará cerca del 90% de la bóveda celeste, en la búsqueda de exoplanetas por el método del tránsito Ricker et al. (2014). En este trabajo se presenta un algoritmo rápido y exclusivo para la detección de tránsitos de exoplanetas en curvas de luz de alta cadencia como las que estará generando el proyecto TAOS II.

El algoritmo que se presenta en este trabajo de tesis consiste principalmente en dos partes: el filtrado de ruido, y la validación de posibles candidatos a tránsitos. Se implementaron 3 métodos para el filtrado de ruido: filtro pasa bajas, promedio móvil y análisis de componentes principales. Para los 3 métodos, se encontró que mejoran notablemente la señal a ruido de las curvas de luz. El método de análisis de componentes principales es el más eficiente y el que tiene un mejor desempeño, mientras que el promediado móvil es el menos eficiente y el método con más error a la hora de ajustar un modelo, esto debido a la distorsión en la curva de luz debido al mismo filtrado.

Después para determinar si un objetivo es un posible candidato a tener un planeta transitando, se comparan las curvas de luz con bases de datos de tránsitos simulados. Aplicando un modelo de regresión, se utiliza como primera aproximación un trapecioide para caracterizar la curva del tránsito. Basados en el parecido entre el modelo y las curvas filtradas, se decide si etiquetar el objetivo como posible candidato a tránsito.

Para evaluar el desempeño de dicho algoritmo se han llevado a cabo temporadas en el OAN-SPM donde se obtuvieron datos con y sin tránsitos de exoplanetas, bajo condiciones de observación muy similares a las que se espera del proyecto TAOS II. Utilizamos APPHi para obtener 13 curvas de luz de ingresos-egresos de tránsitos, de los cuales pudimos detectar exitosamente 5, aunque con diferencias significativas en las observables calculadas.

# CAPÍTULO I: INTRODUCCIÓN

## *I.1 Contexto histórico*

En el pasado se pensaba que el planeta Tierra era el centro del universo, más adelante con el surgimiento de la ciencia como la conocemos hoy en día, se abandonó esta idea basándose en observaciones de los cuerpos celestes en el cielo nocturno. En la era moderna, en 1930 el astrónomo Clyde Tombaugh (1906-1997) descubrió el entonces considerado noveno y más pequeño planeta del sistema solar, Plutón. Para este entonces ya se conocía la complejidad de nuestro sistema solar, compuesto por planetas, satélites naturales, asteroides, cometas y más cuerpos menores. Sin embargo siempre existió la pregunta: *¿Existen otros sistemas solares como el nuestro en el universo?*

En la década de los ochenta, un equipo encabezado por el astrónomo Gordon Walker, planeo utilizar el telescopio CFHT (del inglés Canadian France Hawaii Telescope) de 3.8m volcán Mauna Kea en Hawái, para intentar encontrar la presencia de un planeta alrededor de una estrella lejana (Walker, 2012). En ese entonces ya se conocía que Júpiter completa una órbita alrededor del Sol en aproximadamente 12 años terrestres, y que debido a su gran masa, la interacción gravitacional entre estos dos cuerpos provoca que el Sol se mueva alrededor del baricentro del sistema, con un periodo igual al de Júpiter. La hipótesis del Dr. Walker para la búsqueda de un planeta se basaba en este fenómeno, si era capaz de medir una estrella moviéndose con un periodo orbital similar al de Júpiter, entonces podría asumir que esta estrella poseía un planeta de masa similar a la de Júpiter. Se comenzó con una muestra de 29 estrellas de tipo solar, relativamente cercanas para las cuales se intentó medir su movimiento utilizando la técnica de velocidad radial. Unos años después, se analizaron las 29 estrellas de la muestra inicial sin ningún resultado positivo.

No fue hasta 1995 cuando Michel Mayor y Didier Queloz (Mayor and Queloz, 1995) encontraron una estrella con un movimiento extraño. 51 Pegasi, una estrella de una muestra de 142 que habían estado bajo observación, con una temperatura similar a la del Sol (5768 K), a unos 50 años luz ( $1 \text{ año luz} = 9.461 \times 10^{15} \text{ m}$ ) de distancia. Los datos de la observación revelaron que 51 Pegasi tiene un periodo orbital respecto al baricentro del sistema de 4.2 días, lo cual resultó asombroso y difícil de creer para los autores. El hecho de que un planeta como Júpiter completara una órbita alrededor de una estrella tipo solar en tan solo 4.2 días parecía increíble. Después de corroborar las observaciones, finalmente el 23 de Noviembre de 1995 se publicó el hallazgo de un planeta de una masa similar a la de Júpiter. En 2019 Michel Mayor y Didier Queloz fueron galardonados con el premio Nobel de Física por el descubrimiento del primer planeta alrededor de una estrella tipo solar, lo que actualmente conocemos como exoplaneta.

En los meses siguientes al descubrimiento de 51 Pegasi b en 1995, se encontraron planetas con periodos similares, como los reportados en (Marcy and Butler, 1996) y (Butler and Marcy, 1996). En los años siguientes se siguieron descubriendo planetas en estrellas de tipo solar utilizando esta misma técnica, de hecho, en 2003 el Dr. Walker y un equipo internacional, confirmó la existencia de un planeta del tamaño de Júpiter en la muestra inicial de 29 estrellas que estudió en los 80's, la cual no había sido detectado antes, debido a las limitaciones instrumentales de la época (Hatzes et al., 2003), (Walker, 2012).

Años después, David Charbonneau y Gregory W. Henry estudiaban una de las estrellas de la muestra original del Dr. Mayor, para la cual ya se había confirmado la existencia de un exoplaneta utilizando mediciones de velocidad radial. Sin embargo, estaban en busca de variaciones de brillo debido al paso del planeta frente a la estrella, este fenómeno es conocido como tránsito. En 1999 publicaron el hallazgo del tránsito de HD209458b (Charbonneau et al., 1999), un planeta 60% más grande que Júpiter, con una órbita de 3.5 días, el cual provocó una caída del 1.7% en el brillo de HD209458. Este mismo año se publicó el descubrimiento del primer sistema múltiple descubierto por el método del tránsito, alrededor de *Upsilon Andromedae* (Marcy et al., 1999), 3 planetas de masas similares a la de Júpiter que orbitan una estrella de  $1.28 M_{\odot}$ .

Estos primeros planetas en ser descubiertos tenían algo en común, eran de una masa similar a la de Júpiter con periodos orbitales cortos, los cuales se denominan Júpiteres calientes (*Hot Jupiters*); o tenían órbitas muy excéntricas como el caso de HD222582b (Vogt et al., 2000), ambas características son bastante diferentes a lo que conocemos en nuestro sistema solar, lo cual originaba más preguntas: *¿por qué los sistemas planetarios descubiertos son tan diferentes al nuestro?, y entonces, ¿cómo se formó nuestro Sistema Solar?*, preguntas para las cuáles, aún no tenemos una respuesta clara.

## *I.2 Proyectos enfocados en la búsqueda de exoplanetas por el método de tránsito*

Como hemos comentado, los primeros exoplanetas fueron descubiertos utilizando el método de velocidad radial. Para llevar a cabo una medición de este tipo, se necesita de un espectrógrafo de muy alta resolución y un largo tiempo de monitoreo a una sola estrella, para obtener las mediciones de velocidad que corroboren la existencia de anomalías gravitacionales provocadas por un planeta. Por otra parte, a partir del descubrimiento del tránsito de HD209458b, el método del tránsito se volvió el más exitoso para buscar planetas que orbitan otras estrellas. Una de las ventajas más importantes es que se puede monitorear la intensidad de muchas estrellas al mismo tiempo, como se pretende hacer en el proyecto TAOS II, el cual observará alrededor de 10,000 estrellas simultáneamente.

Desde entonces y hasta la actualidad, se han desarrollado cerca de 50 proyectos enfocados en la búsqueda de exoplanetas, midiendo variaciones periódicas en la luminosidad provocadas por un tránsito exoplanetario. Enseguida se mencionan solo algunos de los proyectos más exitosos.

### *I.2.1 CoRoT*

Para mediados de 2006, todos los planetas descubiertos orbitando estrellas lejanas, eran del tipo *Hot Jupiters*, gigantes gaseosos de más de veinte veces el radio del planeta Tierra ( $>20R_{\oplus}$ ). El 27 de diciembre de 2006, se lanzó el primer satélite enfocado en la



búsqueda de exoplanetas por el método de tránsito, el *CO*nvection *RO*tation and *pl*anetary *TR*ansits (CoRoT), dedicado a la búsqueda de planetas rocosos un poco más parecidos a la Tierra (Aigrain et al., 2007), (Barge et al., 2007), cosa que desde la superficie terrestre no era posible debido a los efectos de distorsión atmosférica (véase sección II.3).

Originalmente CoRoT fue pensado como un proyecto de astrosismología<sup>1</sup>, sin embargo, se acordó el incluir la búsqueda de tránsitos de exoplanetas como su nombre lo indica. En mayo del 2007 se reportó el descubrimiento del primer tránsito de un exoplaneta no conocido, el denominado CoRoT-1b (Barge et al., 2008).

La misión CoRoT estaba equipada con un espejo primario de 0.27 m y una cámara de campo amplio que funcionaba en el espectro visible. El telescopio contenía un deflector de rendimiento extremadamente alto para minimizar la luz directa en el detector. La luz capturada por el telescopio se dirige al plano focal, que consta de cuatro CCD con  $2048 \times 4096$  píxeles dispuestos en un patrón cuadrado con una escala de placa de  $2.32''/\text{pix}$  ( $''/\text{pix} = \text{segundos de arco por pixel}$ ). Dos de los CCD estaban dedicados a la astrosismología, mientras que los dos restantes al programa exoplanetas. Las curvas de luz de las estrellas son producidas por el software a bordo que utiliza fotometría de apertura y posteriormente se transmiten a la Tierra cada 512 segundos. Este instrumento es clasificado como satélite de clase mini (630 kg), que se colocó en una órbita polar con una altitud de 896 km y un período de 6,714 s.

Al concluir el proyecto, se recolectaron 163,000 curvas de luz, con más de 500 candidatos a tránsitos, de los cuales solo 34 (7%) fueron confirmados como planetas (Moutou et al., 2013). CoRoT tuvo contribuciones relevantes como el descubrimiento del primer exoplaneta de tipo rocoso el CoRoT-7b (Léger et al., 2009), el estudio de estructuras internas de estrellas y aportó algunas restricciones para las teorías de formación y evolución de sistemas planetarios.

---

<sup>1</sup> La astrosismología estudia la estructura interna de las estrellas mediante la interpretación de sus espectros como ondas sísmicas.

### I.2.2 HATNet y HATSouth

Los proyectos *Hungarian-made Automated Telescope Network* (HATNet) y *Hungarian-made Automated Telescope Network-South* (HATSouth) conforman una red de 13 telescopios distribuidos en 3 continentes, optimizados para la búsqueda de tránsitos de exoplanetas (Bakos et al., 2002);(Bakos et al., 2004);(Bakos et al., 2013).

Esta red de telescopios intenta detectar y caracterizar planetas extrasolares, utilizando telescopios pequeños totalmente automatizados, constituidos con cuatro telescopios f/2.8 de 0.18 m de diámetro en una montura común que cubren un área de ocho grados cuadrados en el cielo, usando cuatro cámaras con CCD's 4K con iluminación frontal y filtros *Sloan*, para obtener una escala de 3.7 ' /pix (' /pix = minutos de arco por pixel). Como se mencionó anteriormente, las principales ventajas de los estudios de tránsito son los grandes campos que se pueden observar, lo que se traduce en un gran número de estrellas observadas simultáneamente, sin embargo, debido a las características ópticas de los telescopios, sus objetivos se limitan a estrellas brillantes (magnitud  $V \leq 13$ ).

Para HATNet y HATSouth, en total, se han realizado observaciones de seguimiento para 3,200 de los 4,200 candidatos. Se tienen 140 objetos subestelares confirmados de los cuales 128 son exoplanetas, el más reciente publicado en diciembre del 2018 (Bakos et al., 2018). Se concluyó que 2,300 son falsos positivos o falsas alarmas. Los candidatos restantes requieren más observaciones de seguimiento.

### I.2.3 SuperWASP

El proyecto *Wide Angle Search for Planets* (SuperWASP) es una red de observadores internacionales dedicados a la búsqueda de exoplanetas, uno de los proyectos con telescopios terrestres líderes en el área con 185 exoplanetas confirmados (Pollacco et al., 2006). El proyecto utiliza dos telescopios robóticos ubicados en el Observatorio del Roque de los Muchachos en la isla de La Palma en las Islas Canarias (SuperWASP-North) y en el Observatorio Astronómico Sudafricano (SuperWASP-South), lo que permite observaciones en ambos hemisferios. SuperWASP-North y SuperWASP-South comenzaron con sus

actividades en 2003 y 2005 respectivamente, poco tiempo después se reportó el descubrimiento de dos planetas de  $1.33$  y  $1.26 R_{Jup}$  (Cameron et al., 2007).

Tanto *SuperWASP-North* como *SuperWASP-South* poseen monturas ecuatoriales con 8 cámaras de campo amplio. Cada cámara está equipada con un CCD que consta de  $2048 \times 2048$  pixeles, con un tamaño de  $13.5 \mu m$ . El campo de visión de cada cámara cubre un área de 64 grados cuadrados con una escala angular de  $13.7''/\text{pix}$ . Las cámaras observan campos extremadamente amplios, y el conjunto completo de 8 cámaras es capaz de monitorear hasta 480 grados cuadrados en el cielo, siendo considerablemente más grande que otros telescopios convencionales. Durante los primeros 6 meses de observaciones, el SuperWASP-North proporcionó más de 6.7 millones de curvas de luz para estrellas con  $V \sim 7.5 - 15$  mag.

Los datos recopilados por las cámaras son automáticamente procesados por un pipeline, que incluye varios pasos. En el primer paso, el *pipeline* corrige las imágenes eliminando errores causados por el detector o defectos ópticos. Posteriormente, las imágenes se reducen, de forma convencional, al aplicar la calibración de *Bias master*, *Darks* y *Flats*, para cada noche de observación. La fotometría de apertura se utiliza para medir el flujo estelar y construir las curvas de luz de cada fuente individual que se logra identificar en los pasos anteriores. Finalmente, la detección de tránsitos periódicos se realiza aplicando el algoritmo *Box Fitting Algorithm* a las curvas de luz (Kovács et al., 2002), el cual consiste en considerar el cambio en el flujo debido al tránsito como una caja de ancho  $L$ , profundidad  $H$  y un periodo  $P_0$ . Para más detalles sobre la selección de candidatos, ver (Collier Cameron et al., 2006).

#### I.2.4 Telescopio espacial Kepler

La misión Kepler es por mucho, el proyecto más exitoso dedicado a la búsqueda de exoplanetas por el método del tránsito (Borucki et al., 2010). La misión fue diseñada con el objetivo de encontrar planetas rocosos, con masas y tamaños similares a la Tierra en la zona de habitabilidad. La misión se lanzó el 6 de marzo del 2009, tres años después se anunció el descubrimiento del primer planeta rocoso en la zona habitable *Kepler-22b*

(Borucki et al., 2012), un planeta con  $2.4R_{\oplus}$  el cual se denominó como SuperTierra. Conforme avanzaba el proyecto, más descubrimientos importantes se anunciaban, con esto, nos dimos cuenta que los sistemas múltiples y planetas en la zona habitable son más comunes de lo que pensábamos.

La misión estaba diseñada para observar constantemente alrededor de 150,000 estrellas de secuencia principal con magnitudes menores a  $V = 14$  mag en la misma región en el cielo, en la dirección de las constelaciones de *Cygnus*, *Lyra* y *Draco*, una región cerca del plano de la galaxia con alta concentración de estrellas. La sonda espacial *Kepler* está equipada con un telescopio Schmidt 24 de con un espejo primario de 1.4 m de diámetro y una apertura de 0.95m. El detector consiste en una matriz de 42 CCD's que se leen cada 6 segundos para evitar la saturación. Cada CCD tiene un tamaño de 50x25 mm con  $2200 \times 1024$  pixeles, cubriendo un área en el cielo de 105 grados cuadrados.

Los datos obtenidos se almacenaban en la sonda espacial y enviados a la Tierra cada mes. Los datos se recopilan en el Kepler Mission Science Operations Center y eran procesados mediante el pipeline de *Science Processing Pipeline* (Jenkins et al., 2010a), el cual consiste de distintos procesos: calibración de nivel, análisis fotométrico, extracción y corrección de curvas de luz, algoritmo de detección de tránsitos, selección de candidatos y validación de candidatos. Las curvas de luz se extraen mediante fotometría de apertura y posteriormente se corrigen eliminando las firmas correlacionadas con variables instrumentales, como los desplazamientos apuntado y los cambios de foco. Además, se eliminan los valores atípicos (outliers) y las discontinuidades causadas por los cambios de sensibilidad de pixeles. La detección de tránsitos periódicos se lleva a cabo mediante el *Transiting Planet Search (TPS)*, un algoritmo que realiza una estimación el espectro de potencia del ruido de observación en función del tiempo (Jenkins et al., 2010b).

A lo largo de su operación se realizaron muchos descubrimientos importantes, como un sistema planetario orbitando una estrella gigante roja (Charpinet et al., 2011), *Kepler-47* un sistema de dos planetas, uno de ellos en la zona habitable que orbitan un sistema estelar binario (Orosz et al., 2012), *Kepler-90* un sistema de 8 planetas pequeños, todo el sistema es más pequeño que la órbita del planeta Tierra (Shallue and Vanderburg, 2018),

el octavo planeta se confirmó utilizando un algoritmo de Deep Learning desarrollado por miembros de Google, y muchos más. El 30 de octubre del 2018 la sonda se quedó sin combustible y con ello llega el final del proyecto. Después de casi 10 años, se observaron cerca de 530,506 estrellas y con 2662 planetas confirmados, 61 supernovas documentadas, 2878 estrellas binarias eclipsantes confirmadas y alrededor de 3000 publicaciones relacionadas con el proyecto, hacen del telescopio espacial Kepler el proyecto más exitoso en la historia de la búsqueda de nuevos planetas, gracias al cual, ahora sabemos que seguramente que existen más planetas que estrellas en el universo.

### I.2.5 Satélite TESS

El satélite *TESS* del inglés *Transiting Exoplanet Survey Satellite* es el sucesor del telescopio *Kepler*. El objetivo principal de *TESS* es detectar planetas pequeños en estrellas brillantes relativamente cercanas, utilizando el método del tránsito. La primera gran diferencia con *Kepler*, es que *TESS* realizará un monitero de aproximadamente el 90% de la bóveda celeste, con un alcance cerda de  $\sim 60$  pc de distancia (donde  $1\text{pc} = 3.086 \times 10^{16}$ ). La gran ventaja de esto, es que al estudiar principalmente estrellas brillantes, estas pueden ser observadas fácilmente por telescopios en tierra y otros telescopios espaciales. Esto con el fin de facilitar la caracterización y la confirmación de estos exoplanetas, utilizando espectroscopía para medir la velocidad radial de la estrella o buscar medir sus movimientos propios, utilizando datos astrométricos .

Cuenta con 4 cámaras idénticas de campo amplio ( $24^\circ \times 24^\circ$ ) cada una, lo que resulta en un campo de  $24^\circ \times 96^\circ$  (un total de 3200 grados cuadrados) y un radio focal  $f/1.4$ . A su vez, cada cámara está equipada con lentes para prevenir la dispersión de la luz provenientes de la Tierra y la Luna y un arreglo de 4 CCD's de  $4096 \times 4094$  pixeles con una escala angular de 21 .

Los CCD's de las cámaras producen constantemente imágenes de 2 segundos de exposición, las cuales se combinan de 2 diferentes maneras. Primeramente se suman 900 imágenes para simular una exposición continua de 30 minutos del campo completo llamado *Full-Frame Images (FFIs)*, también se suman grupos de 60 imágenes como exposiciones de 2 minutos catalogadas como "*Postage Stamps*", las cuales son preprocesadas y recortadas en secciones de  $10 \times 10$  pixeles centradas en estrellas de interés. (Ricker et al., 2014).

Se espera que el proyecto tenga una duración de 2 años, en el cual se observe completamente el hemisferio sur el primer año y el hemisferio norte en el segundo. Se utiliza un método de apuntado denominado "*stop and stare*", lo que significa que observará un mismo campo completo ( $24^\circ \times 96^\circ$ ) por una duración de 24 días, para después rotar  $27.7^\circ$  hacia el este y repetir el proceso de observación. Con esta estrategia se necesitarán un total de 26 sectores para cubrir el 90 % del cielo (Schliegel, 2017).

*TESS* está enfocado en estrellas de tipo espectral F5 a M5, es decir estrellas relativamente pequeñas en las cuales es más sencillo encontrar planetas pequeños del tipo rocoso (véase sección II.1.1). *CoRoT* y *Kepler* tenían mayor precisión fotométrica capaz de detectar planetas más pequeños, sin embargo, la desventaja es que estos planetas no son viables para su estudio continuo debido al poco brillo de sus estrellas anfitrionas, por lo que sería prácticamente imposible realizar estudios sobre sus propiedades atmosféricas usando espectroscopía (Sullivan et al., 2015).

El satélite fue lanzado el 18 de abril de 2018 y entró en completa operación el 25 de Julio del mismo año. Al momento de la escritura de este trabajo, *TESS* ha confirmado 37 nuevos planetas y se tienen alrededor de 1600 candidatos, entre ellos el primer planeta del tipo rocoso, en la zona de habitabilidad, con un radio de  $1.22R_\oplus$  y una órbita de 37.42 días. (Gilbert et al., 2020)

Se espera que en los 2 años de operación, la sonda *TESS* descubra alrededor de 20,000 planetas, con un estimado de 500 con un radio menor a  $2R_\oplus$ .

### I.3 Proyecto TAOS II

El proyecto *Transneptunian Automated Occultation Survey* (TAOS II) es producto de la colaboración entre instituciones de Taiwán (ASIAA), Estados Unidos (CfA), Canadá (CADAC) México (IA-UNAM). El proyecto contará con 3 telescopios robóticos, con un espejo primario de 1.3 m de diámetro cada uno. Cada telescopio tendrá instalada una cámara enfriada con helio y un detector de 80 Mpx controlado por 10 FPGAs (del inglés *Field Programmable Gate Array*). Con esto se podrán tomar imágenes a 20 imágenes por segundo (fps del inglés *frames per second*) de más de 10,000 estrellas simultáneamente (Lehner et al., 2012).

El proyecto TAOS (2005-2011) es el precursor de TAOS II, el cual contaba con teles-

copios mas pequeños y cámaras menos rápidas. Se encuentra ubicado en el observatorio de Lulin en Taiwán, lugar donde sólo se dispone de menos de 100 noches de observación útiles por año, debido a la calidad del cielo (Lehner et al., 2010).

TAOS II, está ubicado en el Observatorio Astronómico Nacional en San Pedro Mártir (OAN-SPM) en Ensenada B.C., México, donde se tienen mas de 250 noches de observación al año, un cielo más oscuro, menor humedad y mejor *seeing*.

El objetivo principal es detectar ocultaciones de estrellas, por objetos pequeños ( $\sim 1$  km de diámetro) del cinturón de Kuiper. Descubierta en 1992, el cinturón de Kuiper (Luu and Jewitt, 2002) se encuentra más allá de Neptuno, en un rango aproximado de 30 a 50 UA ( $1 \text{ UA} = 149.597 \times 10^6 \text{ km}$ ). Está formado por un gran número de cuerpos del orden de 0.5 - 30 km de radio con una órbita heliocéntrica. El estudio de estos objetos, puede brindarnos la clave para el entendimiento de la formación del sistema solar, así como del origen de los objetos mas lejanos en el mismo, tales como cometas y objetos más allá de Neptuno (TNOs).

Estos eventos son muy raros, la probabilidad de observar una ocultación estelar por objetos del cinturón de Kuiper es de  $10 \times 10^{-4}$  a  $10 \times 10^{-2}$  eventos/estrella/año y tienen una duración muy corta ( $\sim 200$  ms), debido a estos factores, se pretende monitorear una gran cantidad de estrellas simultáneamente a una alta cadencia, para así aumentar significativamente la probabilidad de detección de ocultaciones. Los datos del proyecto generarán una base de datos enorme, alrededor de 5 PB ( $1 \text{ Petabyte} = 1 \times 10^{15} \text{ bytes}$ ), cuya infraestructura ya esta siendo instalada en el Instituto de Astronomía de la UNAM en Ensenada (IA-UNAM).

La base de datos de TAOS II estará a la disposición de la comunidad científica involucrada en el proyecto, y estos podrían ser utilizados para múltiples propósitos, diferentes a los objetivos principales del proyecto. Este trabajo de tesis es uno de esos casos, donde se busca encontrar candidatos a exoplanetas utilizando las curvas de luz que generará el proyecto, implementando un pipeline de procesos para automatizar el análisis y la clasificación de los objetivos. A pesar de las complicaciones, esto puede ser un buen aporte en la comunidad científica que estudia exoplanetas. Los campos de observación de TAOS II serán de 2.3 grados cuadrados cerca de la eclíptica, donde se ha corroborado que no hay traslape con ninguno de los proyectos dedicados a búsqueda de exoplanetas mencionados anteriormente, a excepción de TESS.

### *I.4 Objetivos, hipótesis y estructura de la tesis*

Debido a que la detección de exoplanetas no es el objetivo principal del proyecto, el esquema de observación y la toma de datos presentan algunas complicaciones, ya que están optimizados para detectar eventos rápidos, con una duración típica de una fracción de segundo, en lugar de eventos de largo periodo como un tránsito exoplanetario. Adicionalmente, con el propósito de tener la mejor fotometría posible, se observará con la mínima masa de aire, se restringe la duración de la corrida de observaciones en cada campo a unas 2-3 horas cerca del cenit. Por lo que las herramientas y metodologías convencionales para la identificación de tránsitos, no pueden emplearse exitosamente con los datos de TAOS II. El objetivo es tener una herramienta, capaz de analizar las curvas de luz de la base de datos de TAOS II de manera automática y eficiente, identificando candidatos a tránsitos de exoplanetas, los cuales después se podrán estudiar a mayor profundidad por medio del seguimiento con otros telescopios.

La alta cadencia en las curvas de luz presenta dos principales problemas al querer identificar tránsitos de exoplanetas. Primeramente las curvas de luz contienen una componente de ruido considerable, la cual enmascara cualquier señal producida por un tránsito. Segundo, todas las curvas de luz que se utilizaron en este trabajo tienen una duración de 2 horas, lo cual se traduce en 144,000 muestras, lo cual se vuelve computacionalmente costoso de analizar al momento de tener una gran base de datos ( $\sim 50,000$  curvas de luz por día). Adicionalmente, hay muy pocas probabilidades de poder captar una señal de un tránsito completo en una sola curva de luz. Debido a esto las metodologías explicadas en capítulo II se basan en mejorar la señal a ruido (SNR) utilizando metodologías eficientes y en buscar solo ingresos o egresos de los posibles tránsitos.

En el capítulo III, se presentan las herramientas con las cuales evaluamos el desempeño del algoritmo para identificar posibles candidatos a tránsitos de exoplanetas, tales como datos observacionales, bases de datos simulados y las metodologías para el filtrado de ruido en curvas de luz de alta cadencia.

En el capítulo IV se presentan los resultados obtenidos al analizar los datos simulados y observacionales, siguiendo las metodologías planteadas en el capítulo III.

Finalmente en el capítulo V se presentan las conclusiones del trabajo.



## CAPÍTULO II: MARCO TEÓRICO

### II.1 Señales observadas por tránsitos de exoplanetas

Como se mencionó anteriormente, el método del tránsito es la técnica más popular y exitosa para descubrir nuevos planetas. Un tránsito ocurre cuando la órbita de un exoplaneta está orientada de manera particular con la línea de observación ( $i \sim 90^\circ$ ) y el planeta bloquea parte de la luz de la estrella, al pasar entre esta y el observador lo cual provoca una ligera disminución en su brillo. Como se muestra en la figura 2.1, si la inclinación  $i$  disminuye el tránsito deja de ser visible por el observador, por lo que la probabilidad de observar un tránsito, está en función del ángulo de inclinación con respecto a la perpendicular de la línea de visión.

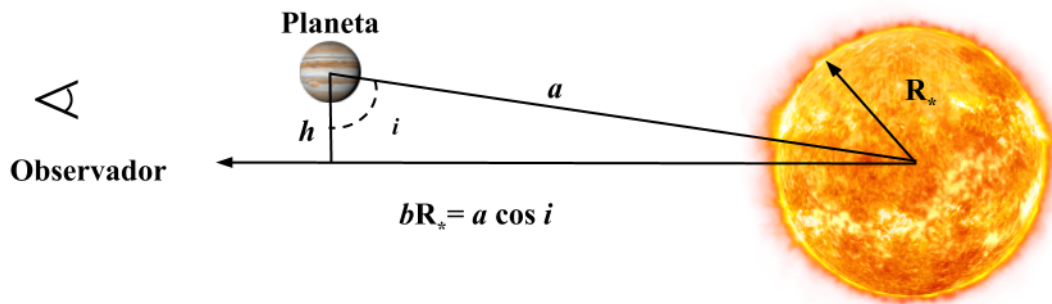


Figura 2.1: Esquema de un tránsito de exoplaneta observable.  $i$  representa la inclinación de la órbita y  $R_*$  es el radio de la estrella.  $b$  es el parámetro de impacto, esta cantidad adimensional es utilizada para resumir la información de la inclinación de la órbita y el semi eje de la órbita  $a$ .

#### II.1.1 Sistema general de ecuaciones

A partir de las señales observadas de tránsitos de exoplanetas, podemos calcular algunas de las propiedades físicas del sistema, teniendo en cuenta las siguientes suposiciones

(para este trabajo):

- La órbita del planeta es circular ( $\epsilon = 0$ ).
- La masa del planeta es despreciable comparada con la de la estrella anfitriona  $M_p \ll M_*$ .
- La relación masa-radio de la estrella es conocida.
- La luz proviene de una sola estrella.
- No se considerará el oscurecimiento al limbo (*Limb Darkening*) véase II.1.2.

Las primeras 3 suposiciones son aceptables para estrellas de secuencia principal, la gran mayoría de los sistemas planetarios cumplen con relaciones similares. Los planetas con periodos orbitales cortos ( $\sim 5$  días) tienen excentricidades cercanas a 0.

Bajo estas aproximaciones, podemos describir la curva de luz de un tránsito a partir de 4 ecuaciones. La primera (2.1) representa la profundidad del tránsito  $\Delta F$ . El flujo total  $F$  de una estrella es proporcional a su superficie, el cambio en el flujo debido al tránsito se traduce en el cociente de las áreas de la estrella y el planeta transitando frente a esta.

$$\frac{\Delta F}{F} = \left( \frac{R_p}{R_*} \right)^2; \quad (2.1)$$

De esta relación podemos claramente ver que si conocemos el radio de la estrella  $R_*$ , podemos entonces calcular el radio del planeta  $R_p$ . También esta es la razón por la cual el proyecto *TESS* se enfoca en estrellas pequeñas, porque de esta manera es más sencillo medir  $\Delta F$  provocados por planetas más pequeños.

La duración total del tránsito  $t_T$  está definida por sus parámetros orbitales y el radio estelar.

$$t_T = \frac{P}{\pi} \arcsin \left( \frac{R_*}{a} \left[ \frac{\left( 1 + \frac{R_p}{R_*} \right)^2 - \left( \frac{a}{R_*} \cos i \right)^2}{1 - \cos^2 i} \right]^{1/2} \right); \quad (2.2)$$

esta relación se puede reducir a:

$$t_T = \frac{P}{\pi} \sqrt{\left(\frac{R_*}{a}\right)^2 - \cos^2 i}; \quad (2.3)$$

sólo considerando que  $a \gg R_* \gg R_p$  (Winn, 2010), donde  $P$  es el periodo orbital dado por la tercera ley de Kepler 2.4,  $a$  es el semi-eje e  $i$  la inclinación de la órbita. Usando  $i$ , podemos definir  $b = \frac{a}{R_*} \cos i$  conocido como parámetro de impacto, el cual es comúnmente más utilizado para cuantificar la inclinación de la órbita. El periodo orbital se define como:

$$P = \sqrt{\frac{4\pi^2 a^3}{GM_*}}; \quad (2.4)$$

donde  $G$  es la constante de la gravitación universal ( $G = 6.674 \times 10^{-11} \frac{Nm^2}{kg^2}$ ), y  $M_*$  la masa de la estrella.

El tiempo total  $t_T$  puede desglosarse en 2 diferentes procesos. El tiempo de que le toma al planeta estar completamente sobre el disco solar, el cual denotaremos como  $t_c$  y la parte “plana” del tránsito  $t_F$ , donde:

$$t_T = 2t_c + t_F; \quad (2.5)$$

Por último, la relación masa-radio, por ejemplo  $R_* = kM_*^x$  ( $x \sim 0.8$  para estrellas de secuencia principal del tipo espectral F-K), y  $k$  una constante el cual su valor depende del tipo estelar de la estrella y su edad (Allen and Cox, 2001).

Utilizando la información contenida en las curvas de luz de tránsitos, podemos calcular cantidades físicas como el radio del planeta  $R_p$  y el periodo orbital  $P$ . El radio del planeta  $R_p$  se obtiene midiendo en la variación del flujo  $\Delta F$  proveniente de la estrella eclipsada (2.1). El periodo orbital  $P$  si se tienen múltiples observaciones del tránsito en diferentes épocas. Para los demás parámetros orbitales se necesita información complementaria de la dinámica de los cuerpos, normalmente obtenida mediante la técnica de velocidad radial. Es por este motivo que para la confirmación de un exoplaneta, este debe ser detectado por un mínimo de 2 técnicas diferentes.

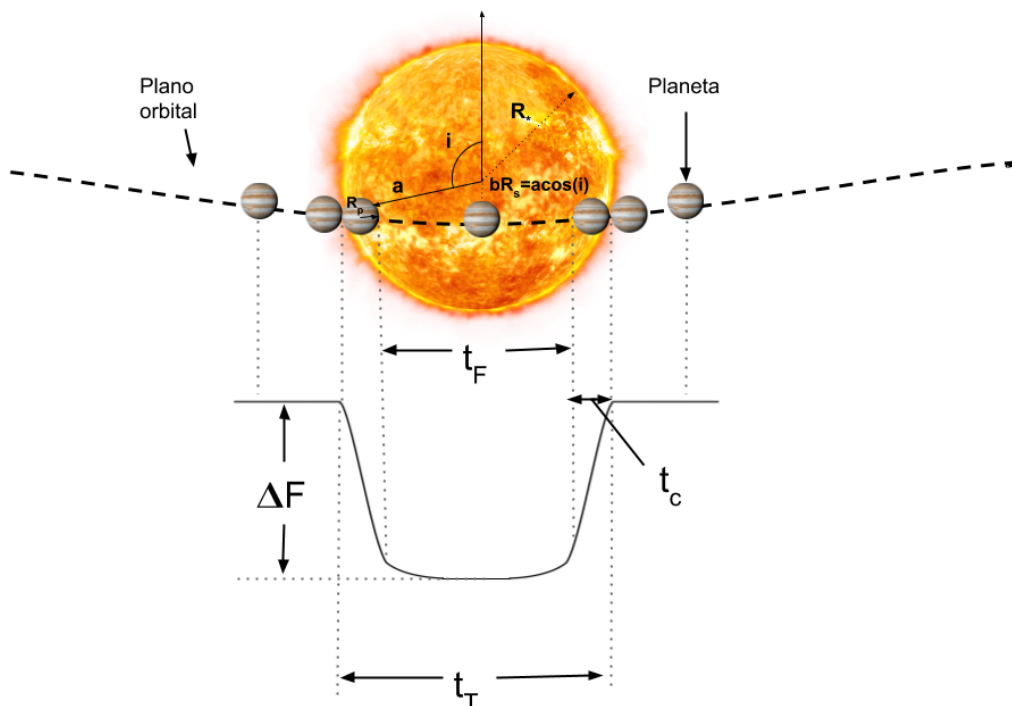


Figura 2.2: Proceso que ocurre durante un tránsito planetario. La curva en la parte inferior representa el flujo de la estrella medido por el observador,  $i$  representa la inclinación de la órbita,  $R_*$  es el radio de la estrella,  $R_p$  el radio del planeta,  $\Delta F$  es el cambio en el flujo visto por el observador,  $t_T$  es la duración total del tránsito y  $b$  el parámetro de impacto.

### II.1.2 Efecto del oscurecimiento del limbo en las curvas de luz

Las ecuaciones de la subsección anterior y sus soluciones no consideran el oscurecimiento del limbo el cual denotaremos como LD (por sus siglas en inglés).

Es posible incorporar el efecto del LD a la descripción matemática del tránsito, incorporando un modelo en el cual el brillo de la estrella varía en función de la distancia a su centro. Esto puede provocar que la profundidad del tránsito  $\Delta F$  se vea afectada, a su vez, se pierde claridad de cuando ocurre el principio/final del eclipse. En el trabajo de (Mandel and Agol, 2002) se muestra la solución analítica de la forma del tránsito, e incorporan el efecto del LD utilizando 3 distintos modelos. En la figura 2.3 podemos observar curvas de luz de tránsitos simulados a partir de las ecuaciones propuestas por (Mandel and Agol,

2002).

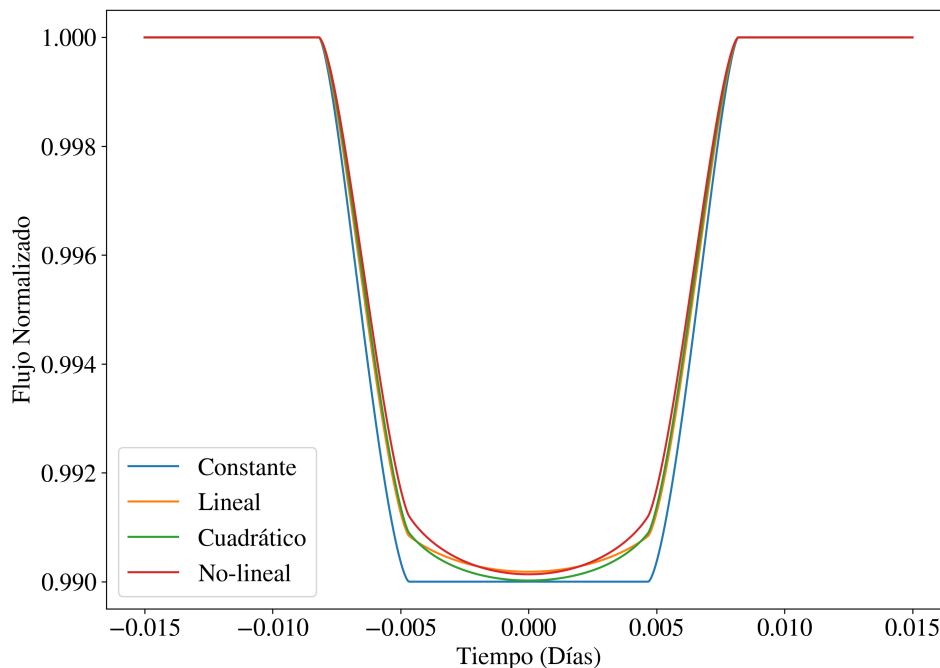


Figura 2.3: Efecto del oscurecimiento del limbo (LD) en curvas de luz de tránsitos simulados. Se muestran 4 modelos distintos, en azul tenemos LD constante, en naranja un modelo lineal, en verde un modelo cuadrático y en rojo un modelo no-lineal.

Para los propósitos de este trabajo, el modelo de LD que se utilice no es importante, debido a que, para la detección del tránsito, solamente buscamos detectar el cambio en el flujo  $\Delta F$  independientemente de la forma de la curva de luz. El modelo de reconocimiento utilizado y descrito más adelante (véase III.6.1) es equivalente a una curva de luz con un LD constante.

### II.1.3 Falsos positivos en la búsqueda de tránsitos

Los grandes proyectos como *TESS* y *Kepler*, observan miles, incluso cientos de miles de estrellas de manera simultánea, esto aumenta significativamente la probabilidad de encontrar sistemas planetarios, cuyo plano orbital se encuentra lo suficientemente alineado

con la línea de visión ( $i \sim 90^\circ$ ), para poder observar un tránsito exoplanetario (véase Figura 2.1). Por lo general, los datos recolectados por estos grandes proyectos, son pre-procesados de manera automática, y se archivan las curvas de luz de interés en las cuales se observan algunas características similares a las que tendría un tránsito real. En el caso de *Kepler*, la estrella se cataloga como *Kepler Object of Interest (KOI)*, sin embargo, un porcentaje de estos objetos de interés no son realmente planetas. Las posibles detecciones fallidas se dividen en dos grupos: falsos positivos y falsas alarmas.

Los falsos positivos ocurren cuando tenemos curvas de luz que parecen ser producidas por un tránsito de exoplanetas pero no lo son. Por ejemplo, para el telescopio espacial *Kepler* se estima que cerca del 10 % de todos los candidatos son falsos positivos (Fressin et al., 2013). Los dos casos más comunes son los siguientes:

- **Estrellas binarias eclipsantes:** son eventos muy similares a los tránsitos, la diferencia es que una estrella es eclipsada por otra estrella compañera. En algunos casos es evidente a simple vista, que la curva de luz no tiene la forma que produciría un planeta. Sin embargo, algunas veces la caída del brillo puede ser muy similar a la producida por un tránsito exoplanetario. Estos falsos positivos son descartados después de un análisis más detallado de la curva de luz, al observar la forma y profundidad del tránsito secundario o analizando las variaciones de fase (Bryson et al., 2017).
- **Binarias eclipsantes o tránsitos de fondo:** ocurren cuando un sistema binario eclipsante o un tránsito sucede en una estrella que ocupan la misma región en el cielo nocturno, es decir objetos que están tan cerca, que no pueden ser resueltos de manera individual. Este fenómeno es un caso particular del primer falso positivo más común.

En la figura 2.4 se muestra un ejemplo de los eventos que pueden dar lugar a falsos positivos. Algunas veces también se observan falsas alarmas, las cuales son curvas de luz con variaciones en el flujo que no son debidas a ningún tipo de tránsito. Pueden ser causadas por variabilidad estelar, condiciones atmosféricas para el caso de telescopios terrestres y/o fallas instrumentales.

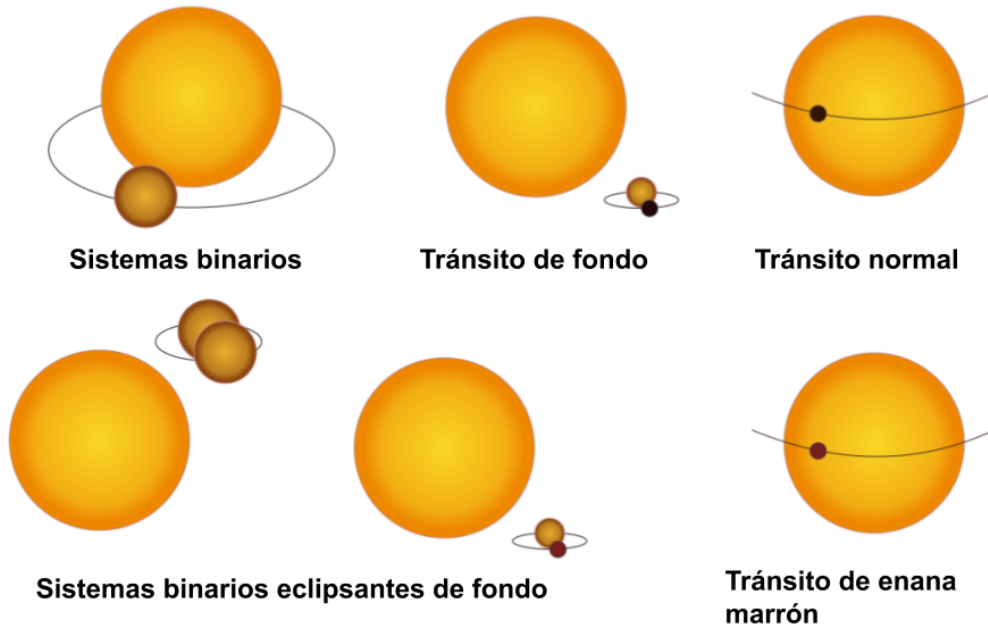


Figura 2.4: Visualización de los distintos casos más comunes de falsos positivos. Imagen tomada de la presentación de Raphaëlle D. Haywood, durante el 2018 Sagan Exoplanet Workshop (liga).

## II.2 Observación convencional de tránsitos de exoplanetas

Los proyectos dedicados a la detección de tránsitos de exoplanetas (véase I.2), forman parte de los más exitosos proyectos en el área. Todos ellos observan estrellas utilizando prácticamente los mismos principios en cuanto a tiempos de exposición de las imágenes obtenidas, y reducción de las imágenes en curvas de luz para su análisis.

Para detectar exitosamente un tránsito exoplanetario, se necesita medir con precisión el flujo proveniente de su estrella anfitriona. Por ejemplo, el tránsito de Júpiter frente a al Sol, provoca una caída en el flujo de  $\sim 1\%$ . Es por esto que se necesitan técnicas fotométricas para aumentar la precisión de la curva de luz. Debido a este mismo motivo, se vuelve complicado detectar planetas más pequeños desde telescopios en Tierra.

Por lo general, se utilizan tiempos de exposición de (30s - 120s), dependiendo de la magnitud de la estrella observada. Cuando el planeta es suficientemente grande, y se tiene una buena SNR (del inglés *Signal to Noise Ratio*) se puede observar a simple vista el tránsito en la curva de luz, en la figura 2.5, se muestra un ejemplo de una curva de luz del exoplaneta XO-3 (Johns-Krull et al., 2008) utilizando 30 segundos de exposición tomada desde el telescopio de 84 cm en el Observatorio Astronómico Nacional (OAN-SPM) el 21 de octubre del 2016, en la cual se puede apreciar el tránsito sin ningún post-procesamiento en la curva de luz.

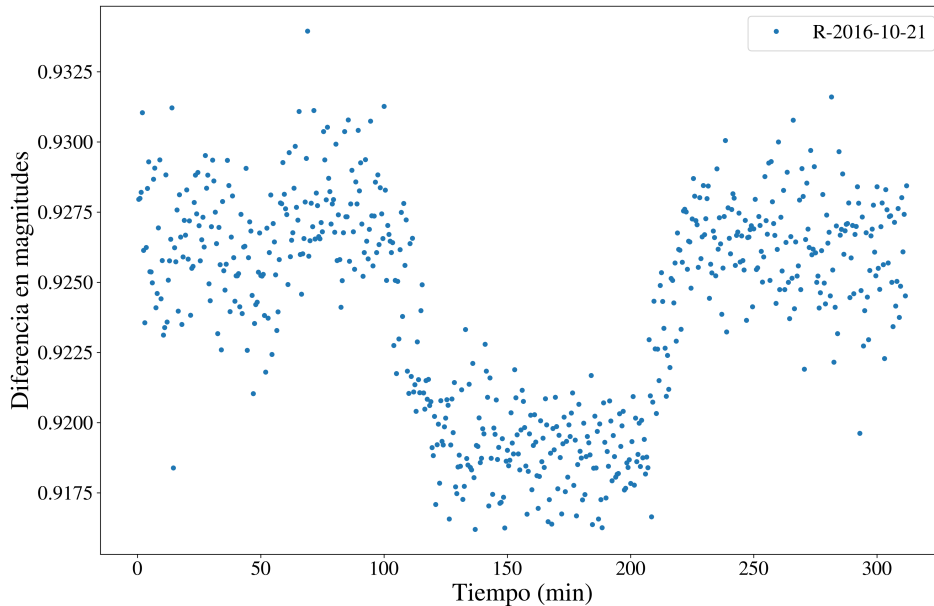


Figura 2.5: Curva de luz de XO-3, tomada con el filtro R de Johnson en el telescopio de 84 cm en el Observatorio Astronómico Nacional (OAN-SPM) el 21 de octubre del 2016.

### II.3 Reducción de imágenes en curvas de luz

Como mencionamos en la sección anterior, necesitamos medir con precisión el flujo  $F_\lambda$  recibido de las estrellas. Es decir, medir la energía por segundo, por unidad de área,



$$F_\lambda = W s^{-1} m^{-2} nm^{-1}. \quad (2.6)$$

Para esto, utilizamos cámaras equipadas con CCD's (*Couple Charge Detectors*) los cuales son un conjunto de “chips”, generalmente de silicio muy sensibles a la luz. Cada CCD está conformado por una matriz de estos “chips”, mejor conocidos como pixeles (para mayor detalle sobre su funcionamiento véase (Mortara and Fowler, 1981)).

Estos dispositivos en combinación con grandes telescopios, nos brindan imágenes de alta calidad. Por lo general, en astronomía se utilizan CCD's monocromáticos debido a que son más sensibles y puede hacerse mejor ciencia con ellos, a diferencia de los CCD's a colores, como las cámaras comerciales para fotografía convencional. Como consecuencia de esto, con la fotometría convencional medimos la intensidad del flujo integrado  $F$  que se define como:

$$F = \int_0^{\text{inf}} F_\lambda d\lambda, \quad (2.7)$$

donde  $F_\lambda$  representa el flujo por unidad de longitud de onda. Sin embargo, el detector es sensible a un rango finito del espectro electromagnético.

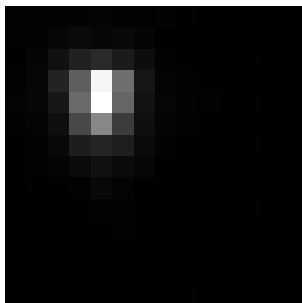


Figura 2.6: Ejemplo de la imagen de una estrella estudiada en este trabajo, obtenida con una cámara rápida IXON en el telescopio de 84 *cm* en el OAN-SPM.

Sin embargo, las imágenes por si solas no son muy útiles cuando se quiere calcular propiedades físicas de los objetos observados, es por eso que necesitamos procesarlas para extraer información de ellas. En este trabajo, estudiamos las estrellas mediante curvas de luz, que son la medida del flujo  $F$  durante intervalos de tiempo regulares, las cuales se obtienen mediante un procesado denominado fotometría. Existen distintas metodologías

dependiendo del objeto que se esté estudiando, en las subsecciones (II.3.2) y (II.3.3) se explican 2 de ellas.

### II.3.1 Preprocesamiento de imágenes

El preprocesado de las imágenes de ciencia nos ayuda a tener mejores resultados después de la fotometría, este proceso limpia las imágenes de variaciones sistemáticas, aberraciones del espejo, ruido electrónico, entre otras. Para esto, es necesario contar con imágenes de calibración comumente llamadas *bias*, *darks* y *flats*, las cuales describiremos a continuación:

- *bias*: estas imágenes son resultado de la lectura del CCD, sin exponer el detector a la radiación. Es decir una imagen tomada con cero segundos de exposición. Estas imágenes contienen información sobre el ruido electrónico y el ruido de lectura del CCD.
- *darks*: estas imágenes resultan de integrar el CCD  $t$  segundos de integración, pero de igual manera que los *bias*, sin exponer el detector a ningún tipo de radiación externa. Se recomienda que los *darks* y las imágenes de ciencia tengan el mismo tiempo de exposición  $t$ .
- *flats*: estas imágenes resultan de exponer todo el detector a una fuente brillante uniforme. Esto nos ayuda a detectar y corregir las irregularidades en las imágenes por el detector mismo, la óptica del telescopio, polvo, etc. Existen diversas maneras de tomar *flats*, para más detalle véase (Chromey, 1996).

Es necesario contar con un conjunto de imágenes de calibración, de 20 a 40, para cada una de las mencionadas anteriormente. Para el tratamiento matemático con estas imágenes, denotaremos a los *bias*, *darks* y *flats* como  $B_t(x, y)$ ,  $D_t(x, y)$  y  $F_t(x, y)$  respectivamente.

El primer paso en el preprocesamiento es obtener las "*master images*" de cada conjunto. El *bias master*  $B_{master}(x, y)$  se obtiene al calcular la mediana por pixel de todo el conjunto de *bias*

$$B_{master}(x, y) = \text{Mdn} [B_t(x, y)]. \quad (2.8)$$

Para calcular el *dark master*  $D_{master}(x, y)$ , restamos el *bias master* al conjunto de *darks* y se calcula la mediana del resultado:

$$D_{master}(x, y) = \text{Mdn} [D_t(x, y) - B_{master}(x, y)]. \quad (2.9)$$

Para calcular el *flat master*  $F_{master}(x, y)$ , primero restamos el *bias master* al conjunto de *flats* y normalizamos a 1 el resultado:

$$F_t^{norm}(x, y) = \text{Norm} [F_t(x, y) - B_{master}(x, y)], \quad (2.10)$$

donde Norm representa:

$$\text{Norm} [I_t(x, y)] = \frac{I_t(x, y)}{\frac{1}{m * n} \sum_{x=0}^m \sum_{y=0}^n I_t(x, y)}. \quad (2.11)$$

Después, calculamos la mediana de  $F_t^{norm}(x, y)$  y normalizamos nuevamente para obtener  $F_{master}(x, y)$ :

$$F_{master}(x, y) = \text{Norm} [\text{Mdn} (F_t^{norm}(x, y))]. \quad (2.12)$$

Finalmente, utilizamos estas *master images* para limpiar las imágenes de ciencia  $O_t(x, y)$ :

$$R_t(x, y) = \frac{O_t(x, y) - B_{master}(x, y) - D_{master}(x, y)}{F_{master}(x, y)}. \quad (2.13)$$

### II.3.2 Fotometría rápida

La fotometría se encarga de integrar el flujo  $F_\lambda$  medido con ayuda del CCD. Con la imagen de una estrella como la que se muestra en la Figura 2.2, se puede obtener una curva de luz mediante la fotometría rápida. Este proceso se realiza tanto como para curvas de luz, como para espectroscopía.

Cada pixel en la imagen tiene asociado un valor comunmente denominado como “cuentas”, este valor representa el número de electrones producidos en ese pixel durante el proceso de captura de fotones. Un CCD ideal, es uno que por cada fotón produce un electrón, los CCD’s de la actualidad están cerca de este límite.

La fotometría rápida consiste en sumar las “cuentas” de cada pixel  $(x_i, y_i)$ , de  $R_t(x, y)$ , donde  $i$  va de 1 hasta  $N$ , donde  $N$  es el número de pixeles en cada eje de la imagen ciencia y  $t$  es el número de imágenes. En este caso consideraremos el caso con imágenes simétricas, donde el número de pixeles en el eje  $x$  ( $N_x$ ), y en el eje  $y$  ( $N_y$ ) son iguales  $N_x = N_y = N$ .

$$F_t = \sum_{x_i}^N \sum_{y_i}^N R_t(x_i, y_i), \quad i = 1, \dots, N; \quad (2.14)$$

La integración de cada imagen produce un solo valor  $F_t$  en la curva de luz (Romanishin, 2006). Por ejemplo, la figura 2.5 fue obtenida mediante la fotometría de 624 imágenes de una región de 7x7 pixeles. Este procesado de las imágenes puede realizarse utilizando distintas metodologías. Como su nombre lo indica, con la fotometría rápida toma poco tiempo generar la curva de luz y es muy útil para análisis in situ. En este trabajo, las curvas de luz fueron obtenidas utilizando la metodología de fotometría de apertura diferencial.

### II.3.1 Fotometría de apertura

Con el fin de mejorar la SNR en la curva de luz resultante de la fotometría, este proceso se mejora realizando lo que se conoce como fotometría de apertura, lo cual consiste en utilizar aperturas circulares para integrar el flujo de la estrella, y además se resta la contribución debido al cielo, aunque en fotometría de alta cadencia, la contribución del cielo es prácticamente nula.

Primeramente definimos 2 máscaras circulares centradas en las coordenadas  $(x_c, y_c)$  del centroide  $C_t$  de la estrella. La apertura de medición  $Area_{M_t}$  mostrada con un área roja en la 2.7 contiene información tanto de la estrella, como del cielo, es decir:

$$M_t = F_t + S_t \quad (2.15)$$

donde  $M_t$  es el flujo integrado dentro de la apertura  $Area_{M_t}$ ,  $F_t$  y  $S_t$  son el flujo en la  $t$ -ésima imagen debido a la apertura estrella y al cielo respectivamente.

El área del anillo formado por los círculos azules, se utiliza para restar la contribución del flujo debido al cielo ( $S_t$ ), siendo  $D_t$  el círculo exterior y  $A_t$  el interior, denotaremos esta área como  $Area_{D_t-A_t}$ , representado por el área azul en la 2.7. Para esto, es necesario

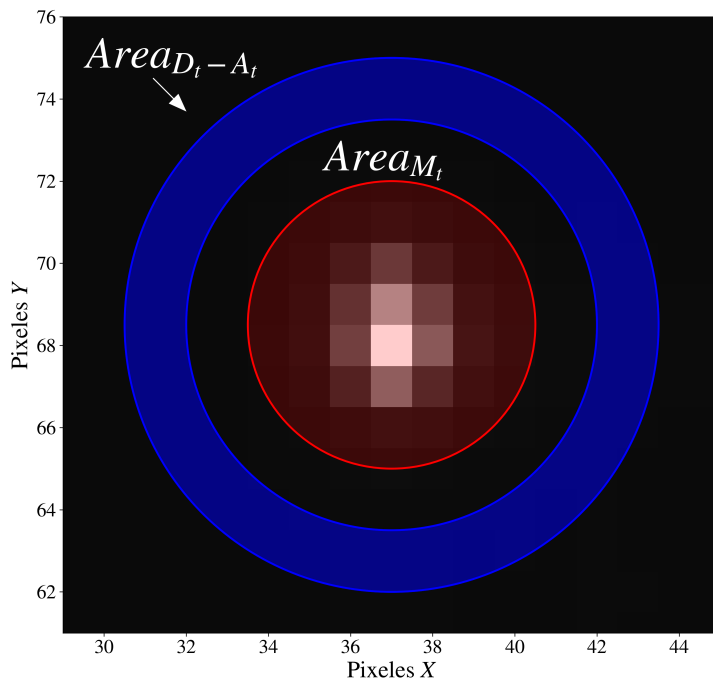


Figura 2.7: Ejemplo de aperturas circulares para realizar fotometría de HAT-P-37b. El anillo azul, formado por las circunferencias  $D_t$  y  $A_t$ , se utiliza para calcular la contribución del flujo debido al cielo  $S_t$ .

primero calcular el flujo del cielo por unidad de área  $S_t^{Area}$ , integrando la contribución de los píxeles dentro del área  $Area_{D_t - A_t}$ .

$$S_t^{Area} = \frac{S_t}{Area_{D_t - A_t}}; \quad (2.16)$$

donde  $S_t$  es el flujo integrado dentro de la región  $Area_{D_t - A_t}$ . El brillo del cielo por unidad de área  $S_t^{Area}$  multiplicado por el área  $Area_{M_t}$  nos da la contribución del cielo dentro de  $Area_{M_t}$ , es decir

$$F_t = F_t + S_t - (S_t^{Area} \times Area_{M_t}); \quad (2.17)$$

Con esta metodología podemos obtener una aproximación del flujo estelar  $F_t$  libre de contribución del cielo. Este proceso se repite para las  $t$  imágenes. Si existe movimiento de la estrella debido a perturbaciones atmosféricas, o mal guiado, el centroide  $C_t$  se recalcula en cada imagen.

### II.3.2 Fotometría diferencial

Si en el campo de observación hay  $N$  estrellas, denotaremos la curva de luz de ellas como  $F^i$  con  $i = 1, \dots, N$ . La curva de luz de una estrella por fotometría diferencial  $F_{Diff}^i$  se calcula utilizando  $F^i$  y el flujo de una estrella compañera  $F^j$ , ambas obtenidas mediante fotometría de apertura.

$$F_{Diff}^i = \frac{F^i}{F^j}; \quad (2.18)$$

Idealmente, estas estrellas deben tener una magnitud similar ( $\pm 10\%$ ) y estar cerca de una de otra ( $< 3$  arcmin) para que la contribución debido a la atmósfera sea comparable.

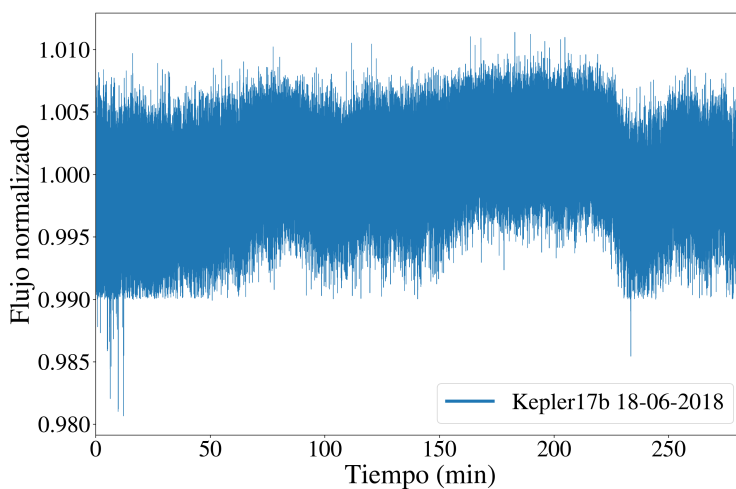


Figura 2.8: Curva de luz de alta cadencia de Kepler-17b obtenida mediante fotometría de apertura. La observación de Kepler-17b se realizó en el sitio 2 de TAOS II, utilizando una cámara *Zyla* de Andor. Esta curva de luz se obtuvo utilizando APPHi véase (sección III.4.1).

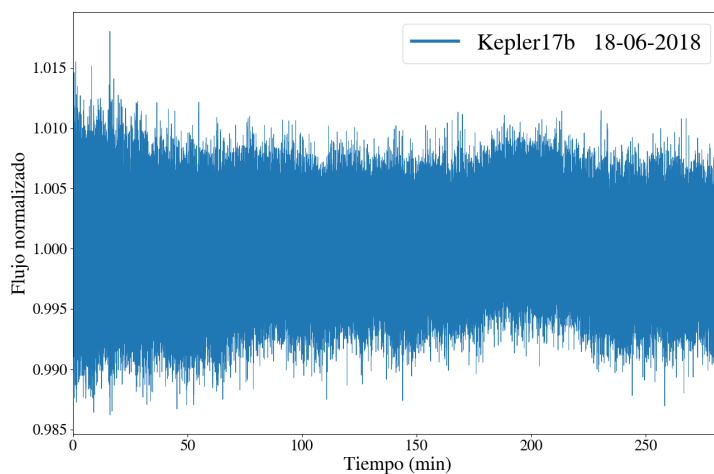


Figura 2.9: Curva de luz de alta cadencia de Kepler-17b obtenida mediante fotometría diferencial. La observación de Kepler-17b se realizó en el sitio 2 de TAOS II, utilizando una cámara *Zyla* de Andor. Esta curva de luz se obtuvo utilizando APPHi véase (sección III.4.1).

En las figuras 2.8 y 2.9 podemos observar la mejora en la curva de luz obtenida mediante fotometría de apertura, gracias a la fotometría diferencial. Esto ocurre debido a que la luz de las estrellas de campo cercanas, sufren deformaciones y variaciones atmosféricas muy similares, gracias a esto, podemos usar el flujo de una estrella para corregir otra curva de luz.

#### II.4 Análisis de ruido en series de tiempo

El análisis de las series de tiempo se basa en suponer que el flujo observado en la curva de luz de la estrella que se está estudiando ( $F_t^{obs}$ ), está compuesto a partir de la contribución de distintos fenómenos:

$$F_t^{obs} = F_t \times R_t; \quad (2.19)$$

donde  $F_t$  es el flujo real de la estrella, sin ningún tipo de contribución debido a la atmósfera o el instrumento de medición. Todas estas contribuciones que añaden ruido al flujo

observado están representadas por  $R_t$ . A su vez,  $R_t$  está conformado por ruido de alta o baja frecuencia.

Las componentes de baja frecuencia se conocen como “tendencias”, se define como un cambio suave del flujo en la curva de luz a largo plazo que afecta a todas las estrellas del campo en forma similar. Generalmente provocadas por condiciones climáticas como la masa de aire, la cual cambia lentamente en unas horas de observación. Otras variables climáticas como densidad del aire, presión y cantidad de agua en la atmósfera cambian más rápidamente, en escalas de tiempo de minutos. Las componentes de alta frecuencia se deben principalmente a ruido electrónico al momento de la lectura del CCD y a la aberraciones del frente de onda causados por la atmósfera. Tiempos de exposición largos ( $T_{exp} \sim 100s$ ) disminuyen la contribución del ruido de lectura, sin embargo en este trabajo nos enfocaremos en estudiar curvas de luz de alta cadencia ( $T_{exp} < 1s$ ), por lo tanto, el detector se lee mas veces por unidad de tiempo, lo cual incrementa en esa misma medida el ruido electrónico (Guerrero et al., 2020).

Estos fenómenos representados en  $R_t$ , afectan la curva de luz creando un ruido aleatorio, el cual puede tener diferentes distribuciones estadísticas como Gaussiana o de Poisson o una convolución de estas distribuciones (Luisier et al., 2010).

Entender el comportamiento de cada componente es importante cuando se desea realizar un análisis numérico, debido a que si se desea simular una curva de luz, esta debe contener la suma de las distintas componentes que conforman una señal real (véase sección II.2). Para el caso de las curvas de luz, estas componentes son catalogadas como tipos de ruido. Como se mencionó anteriormente, el ruido en las curvas de luz, cambia de manera significativa dependiendo de la cadencia de adquisición.

Los distintos tipos de ruido afectan por igual a las imágenes de alta y baja cadencia, y el tratamiento estadístico utilizado para eliminarlo es el mismo en ambos tipos de curvas. Esto puede provocar que la señal de un tránsito en la estrella  $F^t$  se vea ocultada por el ruido  $R_t$  (Pont et al., 2006). Para cuantificar la contribución del ruido, calculamos el coeficiente de señal a ruido (SNR):

$$SNR = \frac{\bar{F}_t^{obs}}{\sigma}; \quad (2.20)$$

donde  $\bar{F}_t^{obs}$  es la media aritmética de la curva de luz y  $\sigma$  su desviación estándar.

Por otra parte, analizar el espectro de frecuencias de la serie de tiempo, nos puede dar



información sobre las distintas componentes de ruido como veremos a continuación.

### II.5 Transformada de Fourier para estudiar series de tiempo

Dentro del estudio de las series de tiempo, una herramienta matemática fundamental es la transformada de Fourier (FT), definida de manera analítica como:

$$\mathcal{F}(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(t)e^{i\omega t} dt; \quad (2.21)$$

donde  $\mathcal{F}(\omega)$  es la transformada de Fourier de  $f(t)$ . Sin embargo, al trabajar con series de tiempo numéricas haremos uso de la transformada de Fourier discreta (DFT):

$$\mathcal{F}_\omega = \sum_{n=0}^{N-1} F_t e^{-\frac{i2\pi}{N}\omega t}; \quad (2.22)$$

de la ec. 2.13 podemos observar que es necesario realizar  $N$  multiplicaciones y  $N$  sumas, lo que resulta en  $N^2$  operaciones. El método más usual de calcular una transformada de Fourier Discreta es utilizar el algoritmo conocido como *Fast Fourier Transform* (FFT). Como su nombre lo indica, es una manera de calcular la transformada más eficientemente, lo cual es útil cuando analizamos una gran cantidad de datos y/o  $N$  es grande. En la FFT se reduce el número de operaciones a  $N \log_2 N$  (Cooley and Tukey, 1965).

Esta transformada integral, nos ayuda a convertir una serie de tiempo  $F_t$ , a una serie de frecuencias  $\mathcal{F}_\omega$ , también llamado *espectro de frecuencias*. Las variaciones sistemáticas en la serie de tiempo  $F_t$  provocan “picos” en el espectro de frecuencias, identificar estas componentes de frecuencia nos ayuda a visualizar las diferentes contribuciones de ruido a la curva de luz.

En la figura 2.10 podemos apreciar una curva de luz de alta cadencia y su respectivo espectro de frecuencias. Como podemos observar, se aprecia una importante componente de bajas frecuencias debido a las tendencias en la curva de luz, provocadas por variables atmosféricas. Es por eso que en este trabajo se modeló el ruido de una manera en la que el ruido simulado tenga componentes que observamos en curvas de luz reales, es decir, que tenga una distribución similar y un espectro de potencias no plano (véase III.2.2), a diferencia de simplemente añadir un ruido con distribución gaussiana (ruido blanco) el

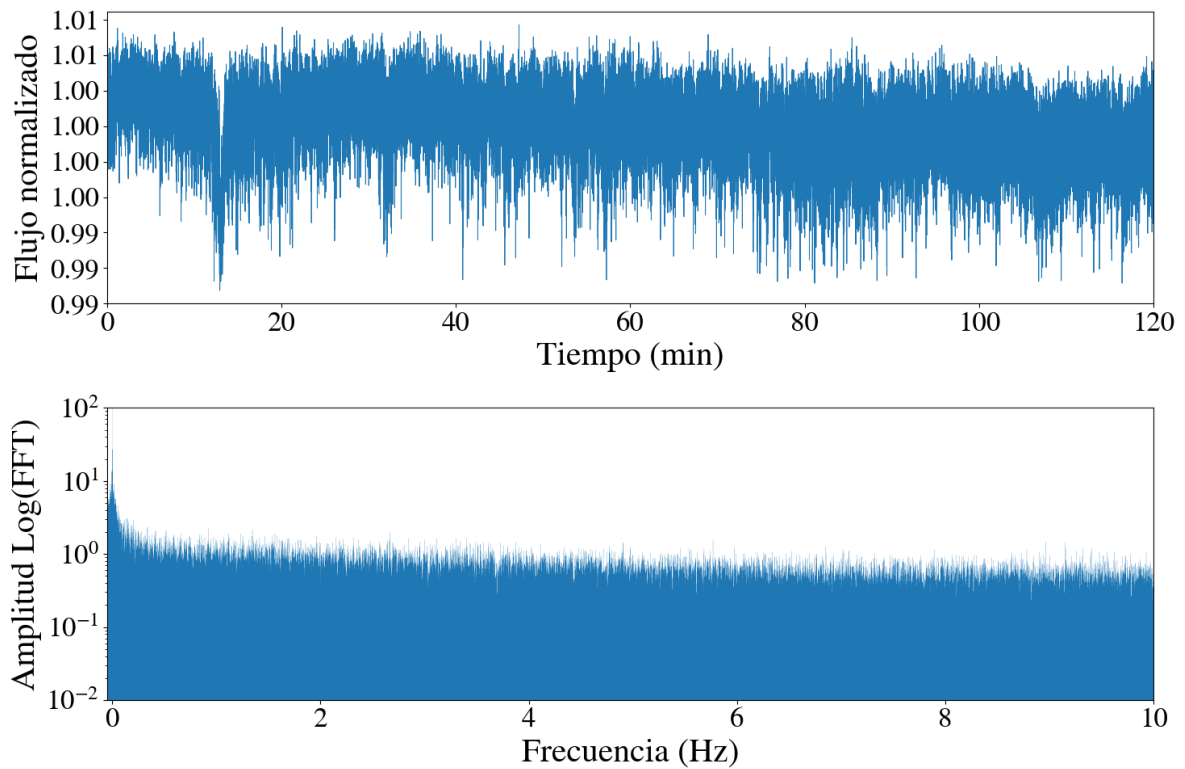


Figura 2.10: Arriba: Curva de luz de alta cadencia de HAT-p-14b, tomada a 20Hz, con el telescopio de 84cm del OAN. Abajo: mostramos su espectro de frecuencias (FFT), con la amplitud en escala logarítmica para mejorar la visualización.

cual tiene un espectro de frecuencias plano.

En este trabajo se utilizó el análisis de Fourier para encontrar las componentes del ruido y limpiar la curva de luz con un filtro pasabajas (LPF de inglés Low-Pass Filter) (Kosarev and Pantos, 1983), (Brault and White, 1971).

## CAPÍTULO III: METODOLOGÍA

En este capítulo se expone la metodología seguida para la obtención de los resultados de este trabajo, además de los conocimientos teóricos básicos y las herramientas numéricas para el análisis de los datos obtenidos.

### *III.1 Datos observacionales, estudios preparatorios para TAOS II*

Si bien la metodología diseñada en este trabajo está pensado para llevarse a cabo con datos del proyecto TAOS II, cuando este en completo funcionamiento. Como parte de este trabajo de tesis, se realizaron observaciones de tránsitos de exoplanetas a una cadencia de 20 fps y escala de placa de 0.61"/pix, es decir, bajo las mismas condiciones que se tendrán en TAOS II (Lehner et al., 2012). Todo esto con el fin, de que cuando el proyecto comience, se tengan las herramientas y metodología para analizar los datos.

Hay que destacar que TAOS II, como ya se mencionó antes, se observarán los campos a  $\pm 1$  hora del cenit, esto para observar con la menor masa de aire posible. Debido a esto será muy difícil observar un tránsito completo, se espera observar cuando mucho la entrada o la salida del exoplaneta frente al disco estelar.

A una cadencia de 20 fps, esto se traduce en curvas con  $N = 144,000$  muestras o datos en cada curva. Además, por campo de observación se espera observar alrededor de 10,000 estrellas en tres telescopios simultáneamente. Por este motivo, los análisis y la metodología presentadas en este trabajo, se enfocan en detectar las firmas de tránsitos de exoplanetas de la manera más eficiente posible desde el punto de vista observacional. Aunque por el esquema de observación que se busca usar, los datos de TAOS II no presentan ventajas fotométricas para la detección de tránsitos por exoplanetas. Solo el hecho de que este proyecto cuente con tres telescopios que estarán observando las mismas estrellas a la vez, constituye una ventaja para descartar falsos positivos, al comparar las curvas de luz y poder indentificar un candidato en utilizando las 3 curvas de luz.



Figura 3.1: Telescopio de 84 cm, en el OAN-SPM, cámara rápida iXon y rueda de filtros MEXMAN.

### III.1.1 Telescopio e instrumentos en la temporada de observación

La mayoría de las observaciones se realizaron en el telescopio de 84 cm en el OAN-SPM (Figura 3.1). Utilizando la rueda de filtros MEXMAN, este instrumento permite intercambiar el filtro que se antepone al detector desde la interfaz de usuario. En total se cuenta con dos ruedas de filtros con 8 posiciones, lo cual resulta en 14 filtros en total (dos posiciones están normalmente vacías).

Como detector se utilizaron dos cámaras en distintas temporadas. La primera de ellas fue del 1 al 6 de agosto del 2016, usando una cámara rápida iXon Ultra 888 de la marca ANDOR, capaz de tomar imágenes de muy corta exposición, con cadencias de hasta 300 imágenes por segundo. Esta cámara fue la única disponible en el OAN-SPM que podía trabajar a 20 fps, hasta que llegaron las cámaras provisionales del proyecto TAOS II. A su vez, las imágenes se tomaron en un *binning*<sup>2</sup> de 3x3, con esto, se logró igualar la escala de placa del telescopio de 84 cm, con los telescopios de TAOS II en donde el campo de visión fue de 3.5 minutos de arco.

La segunda temporada de observación ocurrió del 26 de mayo al 05 de junio del 2017, igualmente en el telescopio de 84 cm. El detector fue una cámara CCD de *frame transfer*

<sup>2</sup> Binning es el proceso de combinar la contribución de varios pixeles adyacentes durante la lectura del CCD.

Nombre	Parámetros del planeta				Parámetros de la estrella			Referencia
	$M_p(M_J)$	$R_p(R_J)$	P (días)	$\Delta F(\%)$	$M_*(M_\odot)$	$m_V$	$T_s$	
HAT-P-14b	$2.2 \pm 0.04$	$1.2 \pm 0.04$	4.62	0.55	$1.38 \pm 0.045$	9.9	F	(Torres et al., 2010)
HAT-P-31b	$2.1^{+0.105}_{-0.07}$	$1.07^{+0.24}_{-0.16}$	5	1.08	$1.21 \pm 0.063$	11.6	-	(Kipping et al., 2011)
HAT-P-32b	$0.86^{+0.105}_{-0.07}$	$1.79 \pm 0.025$	2.15	2.44	$1.12 \pm 0.07$	11.29	F/G	(Hartman et al., 2011)
HAT-P-37b	$1.1 \pm 0.103$	$1.17 \pm 0.07$	2.79	2.04	$0.92 \pm 0.043$	13.2	-	(Bakos et al., 2012)
WASP-48b	$0.98 \pm 0.09$	$1.6 \pm 0.08$	2.14	1.08	$1.19 \pm 0.05$	11.06	-	(Enoch et al., 2011)
WASP-74b	$0.82 \pm 0.014$	$1.4 \pm 0.018$	2.13	1.04	$1.48 \pm 0.12$	9.7	F9	(Hellier et al., 2015)
WASP-3b	$2.6 \pm 0.13$	$1.45 \pm 0.084$	1.84	1.23	$1.24 \pm 0.11$	10.6	F7V	(Pollacco et al., 2008)
Kepler 17b	$2.45 \pm 0.014$	$1.31 \pm 0.018$	1.48	2.13	$1.16 \pm 0.06$	14	G2V	(Désert et al., 2011)

Tabla 3.1: Resumen de propiedades físicas de los objetos observados en distintas temporadas en el OAN-SPM.  $M_p$  es la masa del planeta en unidades de masas de Júpiter,  $R_p$  es el radio del planeta en radios de Júpiter, P es el periodo orbital en días,  $\Delta F$  es el cambio en el flujo de la estrella debido al tránsito en porcentaje,  $M_*$  es la masa de la estrella anfitriona en unidades de masas solares,  $m_V$  y  $T_s$  son la magnitud aparente y el tipo espectral de la estrella respectivamente.

de la marca *Spectral Instruments*. El campo de la cámara fue disminuido para alcanzar la cadencia necesaria (20 fps), al igual el *binning* de las imágenes variaba dependiendo de la magnitud de la estrella.

El 18 de junio del 2018 se obtuvo una curva de luz de alta cadencia (20 fps) de Kepler-17b en tránsito. Se utilizó el telescopio del sitio 2 de TAOS II, en donde se instaló una cámara provisional *Zyla* de Andor.

### III.1.2 Descripción de objetos observados

En la tabla 3.1 se muestra un resumen de los parámetros de los sistemas observados en las múltiples temporadas, algunos objetivos fueron observados en varias ocasiones y bajo distintas condiciones atmosféricas.

### III.1.3 Descripción de datos, influencia del seeing y ruido sistemático

Debido al corto tiempo de exposición, el número de cuentas por estrella es pequeño, por lo cual, el ruido electrónico se vuelve un factor importante. Por otro lado, las perturbaciones atmosféricas afectan el *seeing* que es una medida de la estabilidad de la atmósfera. Esto afecta a las imágenes de alta cadencia de dos maneras: 1) cambiando de posición la estrella en el detector o deformándola, lo cual se convierte en una fuerte contribución al

ruido cuando se lleva a cabo la fotometría de apertura (Guerrero et al., 2020). El clima durante las temporadas fue bueno (sin nubosidad visible) a tal punto de que no hubo contratiempos en las observaciones, aunque se presentan ligeras variaciones en la posición de la estrella de toma a toma. Estas variaciones se ven reflejadas a la hora de obtener la curva de luz de las imágenes. La cantidad de ruido se puede cuantificar calculando la SNR (2.11), la cual está en función de la magnitud aparente de la estrella.

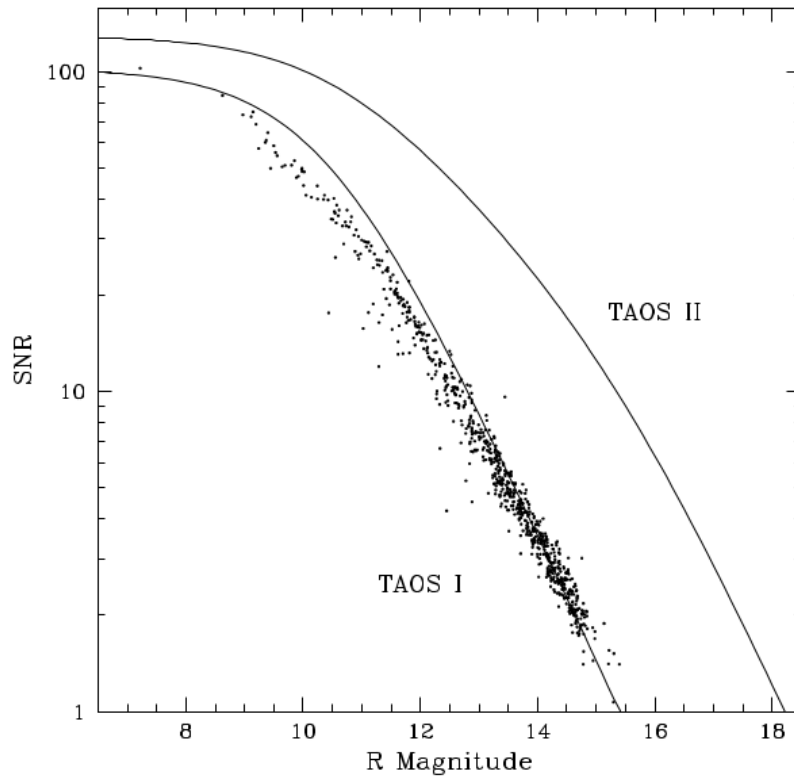


Figura 3.2: SNR en función de la magnitud en el filtro R. Los puntos representan la SNR medida utilizando curvas de luz obtenidas en TAOS I. Las líneas sólidas representan la predicción de SNR utilizando ruido con distribución de Poisson, ruido del cielo y ruido de lectura. Gráfica tomada de (Lehner et al., 2016).

En la tabla 3.2, se muestra un resumen de las curvas de luz obtenidas mediante las diferentes temporadas descritas en III.1.1.

Utilizamos fotometría diferencial para reducir los ruidos sistemáticos, sin embargo, no todos los campos observados contaban con buenas estrellas de comparación, por lo que en algunos casos se utilizó simplemente fotometría de apertura. Como se aprecia en las

Fecha	Nombre	Detector	Fotometría
06/Ago/2016	WASP-74b <sup>1</sup>	iXon Ultra 888	Diferencial
06/Ago/2016	WASP-74b <sup>2</sup>	iXon Ultra 888	Diferencial
06/Ago/2016	HAT-P-32b	iXon Ultra 888	Apertura
07/Ago/2016	HAT-P-37 <sup>1</sup>	iXon Ultra 888	Apertura
07/Ago/2016	HAT-P-37 <sup>2</sup>	iXon Ultra 888	Apertura
28/May/2017	HAT-P-14b	FrameTransfer	Apertura
29/May/2017	WASP-74b	FrameTransfer	Diferencial
31/May/2017	WASP-48b	FrameTransfer	Diferencial
31/May/2017	WASP-74b	FrameTransfer	Diferencial
02/Jun/2017	WASP-48b <sup>1</sup>	FrameTransfer	Diferencial
02/Jun/2017	WASP-48b <sup>2</sup>	FrameTransfer	Diferencial
05/Jun/2017	HAT-P-31b	FrameTransfer	Diferencial
08/Jun/2018	Kepler-17b	Zyla	Diferencial

Tabla 3.2: Curvas de luz con tránsitos de exoplanetas obtenidas mediante las diferentes temporadas descritas en III.1.1. Todas las curvas de luz tienen una duración de dos horas y fueron obtenidas en el telescopio de 84cm en el OAN-SPM. Solamente la observación de Kepler-17b se llevo a cabo en el sitio dos de TAOS II. Los superíndices (<sup>1</sup> y <sup>2</sup>) indican que se observó el tránsito completo, pero la curva de luz se dividió en segmentos de dos horas cada uno (ingreso y egreso del planeta respectivamente), los cuales se analizaron de manera independiente.

figuras 2.8 y 2.9, las curvas de luz resultantes están dominadas por ruido. El propósito de este trabajo de tesis es idear metodologías para caracterizar el ruido y removerlo de las curvas de luz para poder detectar tránsitos de exoplanetas, es decir, detectar variaciones en el flujo de ( $\sim 1 - 2\%$ ).

### III.2 Análisis de datos observacionales

A continuación se plantea la metodología seguida para procesar los datos observacionales resultado de las múltiples temporadas en el OAN-SPM, así como las simulaciones.

#### III.2.1 Fotometría con APPHi

APPHi del inglés *Automated Photometry Pipeline for High Cadence, Large Volume Data* es un código desarrollado en C, el cual consta de dos principales programas, *APPHi*



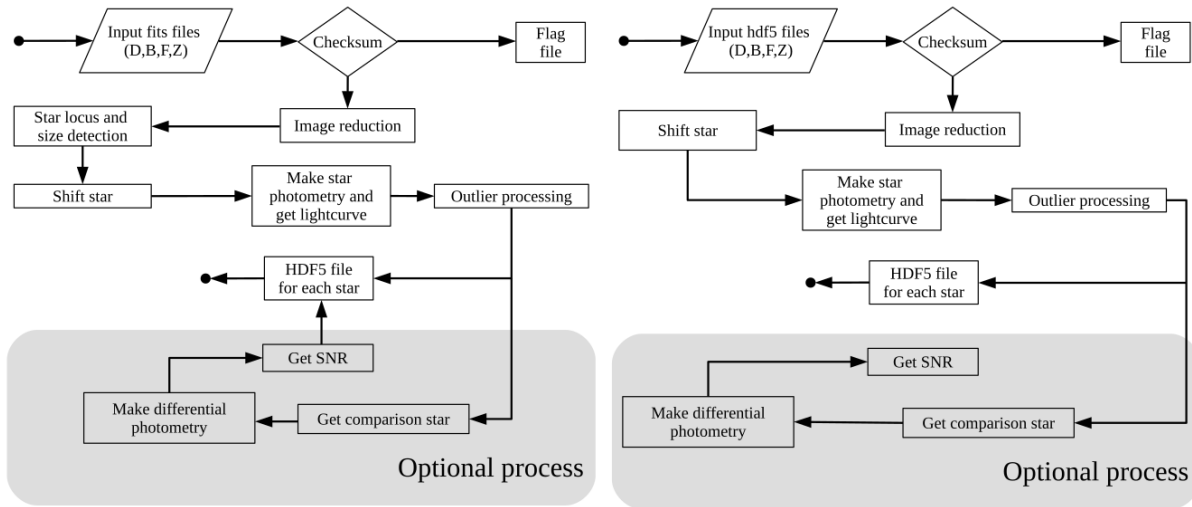


Figura 3.3: Diagrama de flujo para ambos programas (*APPHi field* y *APPHi star*), donde D, B, F y Z representan los darks, bias, flats e imágenes de estudio respectivamente. Imágen tomada de (Sánchez et al., 2019).

*field* y *APPHi star* los cuales funcionan de acuerdo al diagrama de flujo mostrado en la figura 3.3. Este pipeline fotométrico se utilizará como herramienta para procesar el gran volumen de datos que va generar el proyecto TAOS II (Lehner et al., 2016). Este código consta de 3 partes principales, preprocesamiento de las imágenes, localización de las estrellas en el campo de observación y la fotometría para generar las curvas de luz.

Primero las imágenes son preprocesadas de la manera convencional, utilizando imágenes de calibración (bias, flats y darks) como se describe en (Romanishin, 2006). Después de limpiar las imágenes, APPHi localiza de manera automática los centroides correspondientes para cada estrella en el campo de observación usando el método del centro de masa (Chong and Zak, 2004). Este centroide se recalcula para cada imagen, para mejorar la curva de luz. Este proceso iterativo disminuye los efectos telúricos producidos por la atmósfera y el movimiento mecánico del instrumento, los cuales se vuelven significativos a altas cadencias.

Este proceso resulta en cubos de imágenes centradas por estrella, a partir de los cuales se realiza la fotometría de apertura. Este proceso sigue la metodología descrita en la subsección II.3.1 y en (Romanishin, 2006).

APPHi realiza de manera automática otros procesos posteriores a la fotometría para

mejorar la SNR. Uno de ellos es la fotometría diferencial, utilizando curvas de luz de estrellas cercanas como se explicó en la subsección II.3.2. Otro método que se aplica directamente a las imágenes es la técnica *Shift & Add* (Klückers et al., 1996).

Para ejecutar APPHi desde la línea de comando se utiliza:

```
APPHi_field bias.txt darks.txt flats.txt night.txt
```

donde cada uno de los archivos `.txt` contiene la lista de imágenes correspondientes en formato FITS. El resultado del procesado de las imágenes es un archivo HDF5 por cada estrella en el campo de observación. Este archivo contiene las imágenes de la estrella, las coordenadas del centroide para cada imagen, la curva de luz obtenida por fotometría de apertura y fotometría diferencial (Sánchez et al., 2019). Todas las curvas de luz analizadas en este trabajo de tesis fueron obtenidas utilizando APPHi, el cual se encuentra disponible en <http://taos2.astrosen.unam.mx/APPHi.html>.

### III.3 Análisis de ruido en curvas de luz de alta cadencia

A continuación se muestra la metodología para la caracterización del ruido, a partir de las curvas de luz resultantes de las temporadas de observación.

#### III.3.1 Distribución temporal del ruido

El primer paso para la caracterización del ruido, fue saber que tipo de distribuciones podemos encontrar en los datos. En esta subsección analizaremos el ejemplo de la curva de luz de WASP-74b (ver Tabla 3.1 para más detalle) calculada usando fotometría diferencial de las imágenes obtenidas el 27 de mayo del 2017. La metodología explicada a continuación se utilizó para caracterizar el ruido en todas las curvas de luz.

$$G(x) = Ae^{-\frac{(x - \mu)^2}{2\sigma^2}}; \quad (3.1)$$

Un ruido con una distribución gaussiana 3.1 es conocido como ruido blanco. Se puede apreciar un ejemplo en la figura 3.4.

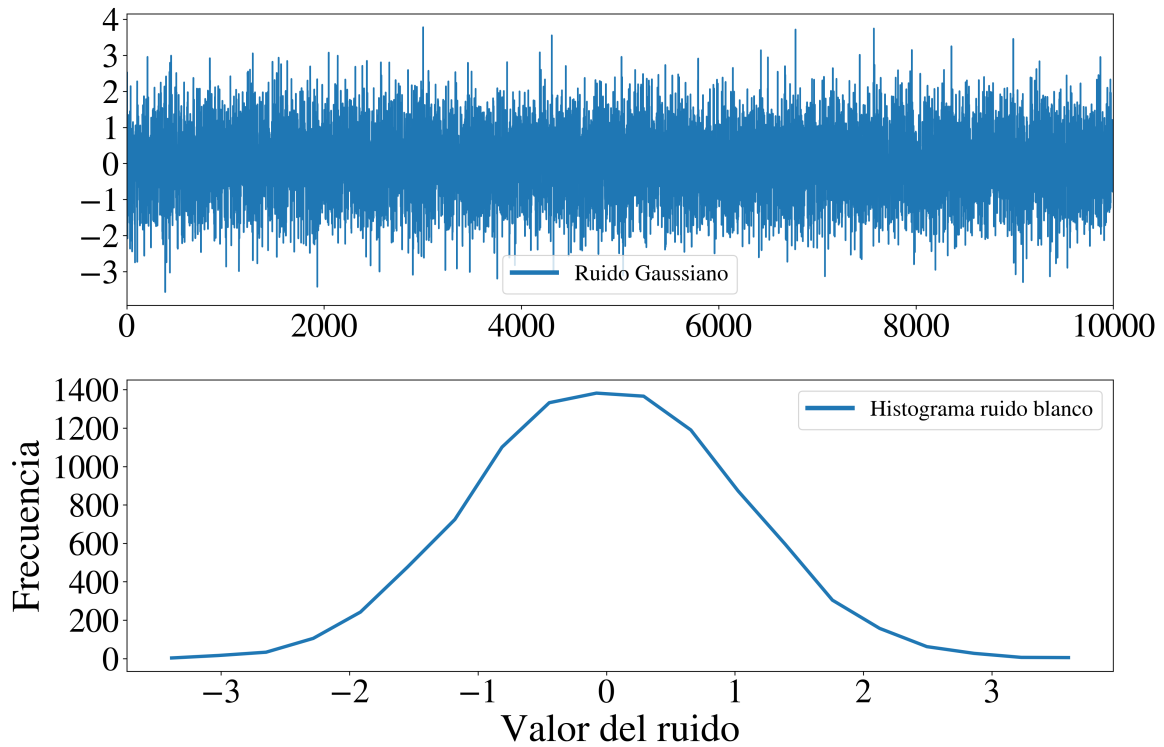


Figura 3.4: Arriba: Ejemplo de ruido blanco (gaussiano). Abajo: Histograma del ruido generado.

Esta es la manera mas común y simple de modelar el ruido, sin embargo, al analizar la curva de luz se encontraron 2 componentes gaussianas que conforman el ruido observado en la curva de luz de WASP-74b que se aprecia en la figura 3.7.

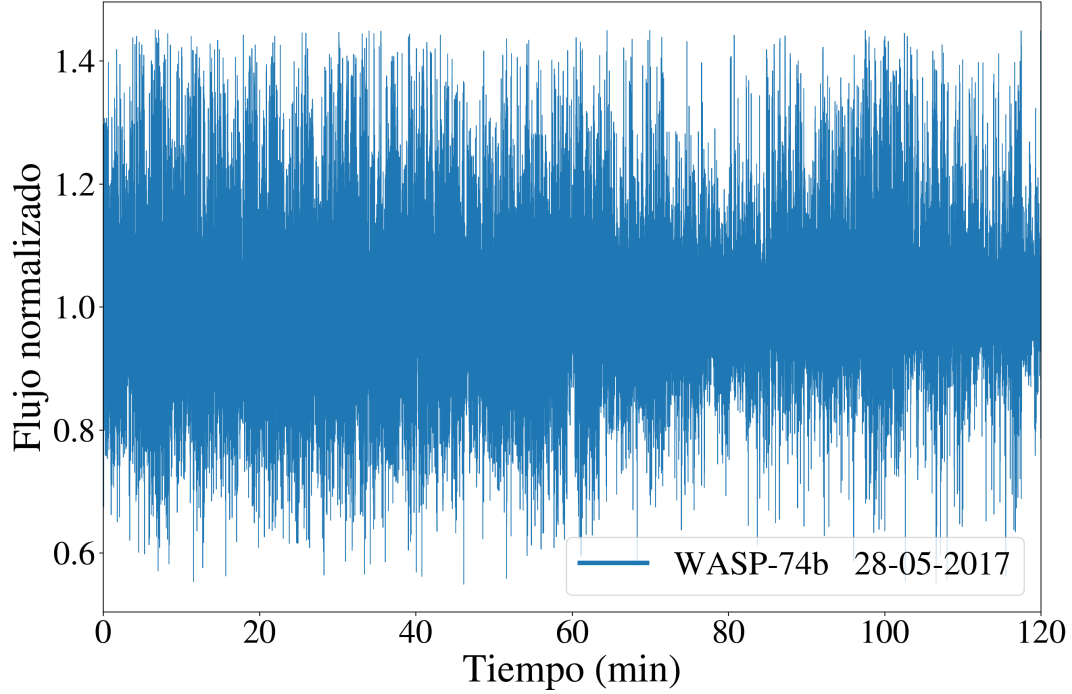


Figura 3.5: Curva de luz de WASP-74b, calculada mediante fotometría diferencial usando APPhI.

Esta misma distribución se encontró en diferentes curvas de luz, por lo que se ajustó un modelo de ruido con dos componentes gaussianas (3.2) a las distribuciones temporales de las curvas de luz.

$$m(t) = A_1 e^{-\frac{(t - \mu_1)^2}{2\sigma_1^2}} + A_2 e^{-\frac{(t - \mu_2)^2}{2\sigma_2^2}} = G_1(t) + G_2(t); \quad (3.2)$$

El resultado del ajuste de  $m(t)$  al histograma de la curva de WASP-74b se muestra en la figura 3.6.

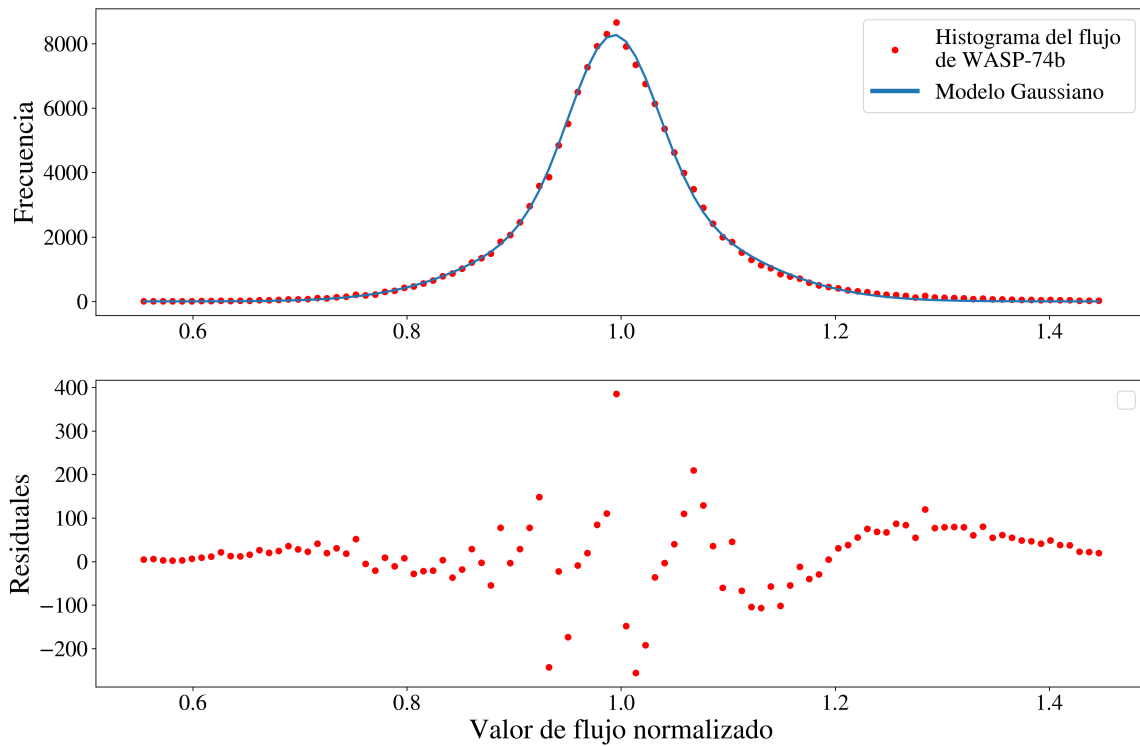


Figura 3.6: Arriba: Los puntos rojos son la distribución temporal de la curva de luz de WASP-74b. La línea azul sólida, es  $m(t)$  resultado de un ajuste no lineal utilizando el algoritmo de Levenberg-Marquardt. Abajo: se aprecia los residuales de la diferencia entre la distribución y el modelo.

La contribución de cada componente  $m(t)$  se aprecia en la figura 3.7. De este gráfico podemos intuir que existen 2 componentes que contribuyen al ruido final en la curva de luz, sin embargo, la fuente de estas componentes no es de interés para los propósitos de este trabajo.

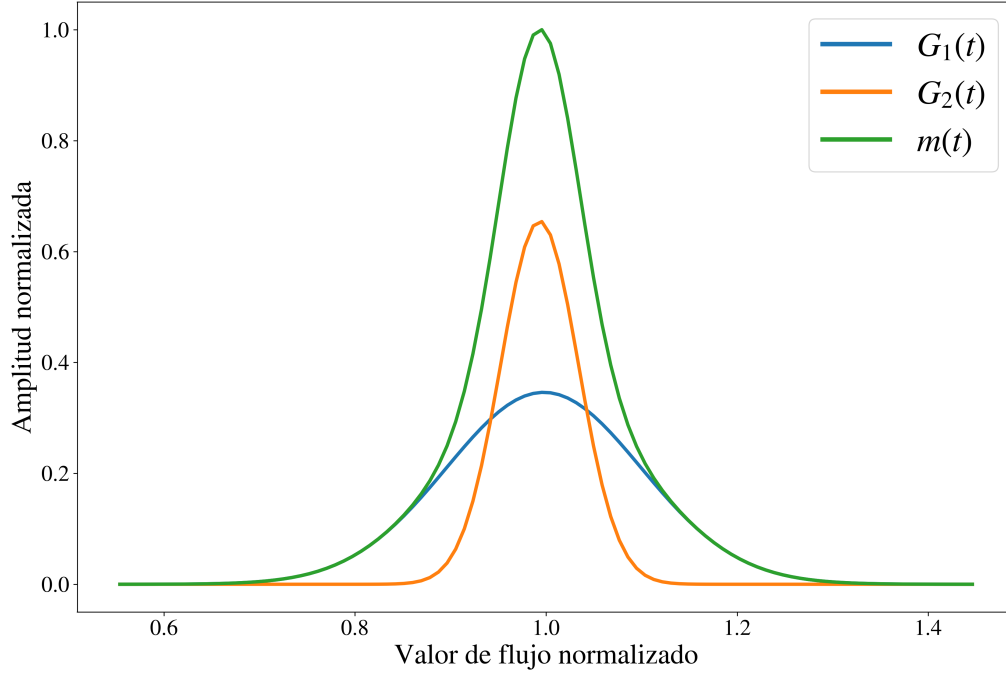


Figura 3.7: Se muestra la contribución de las componentes  $G_1$  y  $G_2$  y la suma  $m(t)$ .

Ajustando  $m(t)$  a todas las curvas de luz, se calcularon los valores extremos o cotas para la amplitud ( $A_1, A_2$ ), las medias ( $\mu_1, \mu_2$ ) y las desviaciones ( $\sigma_1, \sigma_2$ ) de la función de distribución  $m(t)$ , para el momento en que estemos modelando el ruido (véase III.3.3).

### III.3.2 Espectro de frecuencias del ruido

El ruido blanco como el de la figura 3.4 tiene un espectro de frecuencias plano, incluso la suma de 2 componentes de ruido gaussiano sigue teniendo un espectro de potencias plano como se aprecia en la figura 3.8.

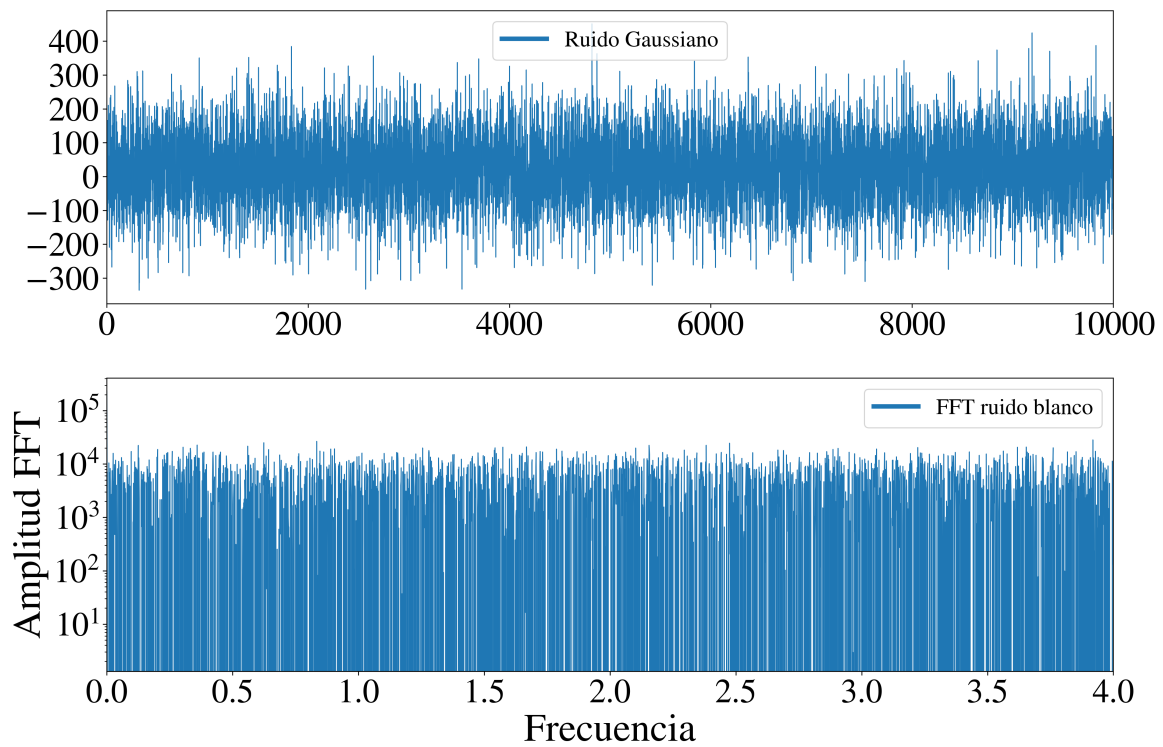


Figura 3.8: Arriba: se muestra una curva de ruido blanco con dos componentes  $G_1$  y  $G_2$ . Abajo: el espectro de frecuencias de la curva de luz, el cual permanece plano sin importar el número de componentes gaussianas.

Sin embargo, al analizar las curvas de luz se encontró que un espectro de frecuencias plano, es solo un caso muy particular, solo 2 de las 13 curvas de luz en 3.2 presentaron un espectro plano. Todas las otras curvas tienen una significativa componente de bajas frecuencias, así como una ligera pendiente negativa, disminuyendo hacia las frecuencias altas como se aprecia en la figura 3.8.

Entonces si queremos simular ruido, que posea características que se aprecian en una curva de luz real, necesitamos más que una simple distribución gaussiana, por lo que esta distribución temporal es solo el punto de partida para generar un ruido sintético que se apegue a las observaciones.

Para añadir estas contribuciones de baja frecuencia, se modeló una envolvente exponencial de la forma:

$$E(f) = C + a_1 e^{-\frac{f}{b_1}} + a_2 e^{-\frac{f}{b_2}}; \quad (3.3)$$

la función envolvente 3.3, está compuesta por dos componentes que decaen de manera exponencial hacia las altas frecuencias. Estas componentes son provocadas por fenómenos de baja frecuencia en la curva de luz, como tendencias o un posible tránsito. La envolvente tiene una asíntota horizontal en  $C$ , que representa el valor mínimo del espectro de frecuencias observado.

Utilizando esta envolvente, se altera el espectro de frecuencias de acuerdo a lo analizado en las curvas de luz de las observaciones. De manera similar a la metodología implementada para las distribuciones temporales  $m_t$  (véase III.3.1), se calcularon los parámetros de  $E(f)$  para todas las curvas de luz y se obtuvieron restricciones para la asíntota  $C$ , las amplitudes  $a_1$  y  $a_2$  y los factores  $b_1$  y  $b_2$  para la posterior simulación de ruido (véase III.3.3).

En la figura 3.9, podemos observar la envolvente exponencial  $E(f)$ , del espectro de potencias de la curva de luz de WASP-74b. Como se puede apreciar en la parte superior (temporal), esta curva contiene una ligera tendencia, y variaciones de flujo en el rango de 1-3 minutos. Estas variaciones contribuyen a que la amplitud del espectro en bajas frecuencias sea significativamente mayor que en altas frecuencias, esto se ve reflejado en los parámetros  $C$ ,  $a_1$  y  $a_2$ . La línea roja es la contribución de un tránsito al espectro, esta se encuentra en la región de bajas frecuencias, alrededor de  $f < 0.01$ . En la tabla 3.3 se muestran los valores de los parámetros de  $E(f)$  obtenidos de los datos observacionales.

Utilizando los resultados obtenidos y siguiendo la metodología de plateada en (III.3.1 y III.3.2) se implementó un algoritmo para generar ruido con características similares, que podemos observar y medir en curvas de luz reales. Así, nuestras simulaciones, tendrán más validez, que en lugar de usar un simple ruido gaussiano.

### III.3.3 Simulación de ruido

Primeramente, para simular un ruido  $R(t)$ , dada una función de distribución  $m(t)$ , por ejemplo 3.2, calculamos la función de distribución acumulativa:

$$M(t) = \int m(t) dt; \quad (3.4)$$



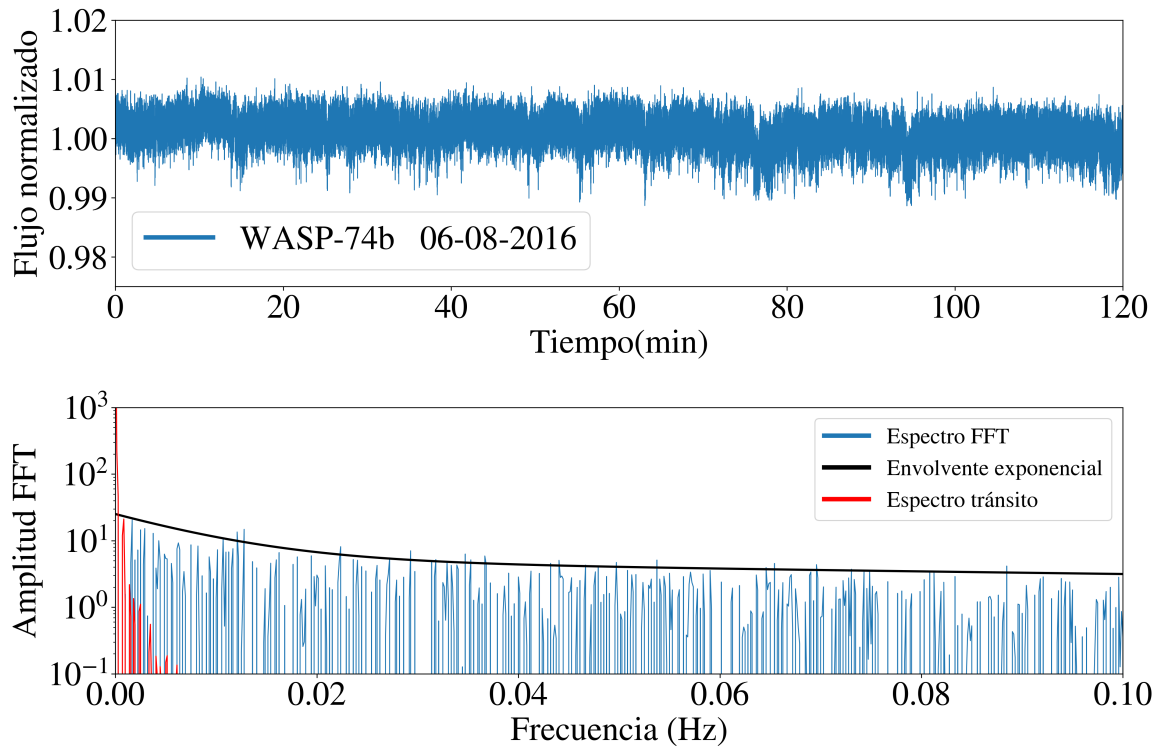


Figura 3.9: Arriba: se muestra una curva de luz de WASP-74b. Abajo: la zona de bajas frecuencias del espectro de la curva de luz, la línea negra es el ajuste de la envolvente exponencial  $E(f)$  del espectro. La línea roja indica el espectro del tránsito inmerso en la curva de luz.

Fecha	Nombre	$C$	$a_1$	$b_1$	$a_2$	$b_2$
06/Ago/2016	WASP-74b	5	25	0.155	105	.0089
06/Ago/2016	HAT-P-32b	5.56	15	0.095	31.85	.0079
07/Ago/2016	HAT-P-37	43	0	-	0	-
28/May/2017	HAT-P-14b	3	10	0.25	51.2	0.0049
29/May/2017	WASP-74b	5	15	2.55	100	0.02
31/May/2017	WASP-48b	10	15	0.085	120	0.01
31/May/2017	WASP-74b	5	5	2.50	26	0.06
02/Jun/2017	WASP-48b	4	15	0.043	41	0.0035
05/Jun/2017	HAT-P-31b	3.5	11.6	0.277	85	0.0083
08/Jun/2018	Kepler-17b	10	30	0.155	180	0.0029

Tabla 3.3: Resumen de los parámetros  $(C, a_1, b_1, a_2, b_2)$  calculados al ajustar la envolvente 3.3 al espectro de potencias de las curvas de luz de tránsitos de exoplanetas obtenidas mediante múltiples temporadas en el OAN-SPM.

y después normalizamos  $M(t)$ :

$$M_{norm}(t) = \frac{M(t)}{\text{MAX}(M(t))}. \quad (3.5)$$

Para generar el vector de ruido, se realiza un muestreo de puntos aleatorios  $t_i$  con distribución uniforme  $\mathcal{U}(a, b)$ , evaluados en la función inversa de la distribución acumulada normalizada, denotada como:  $M_{norm}^{-1}$ . Resumiendo, a continuación se muestran los pasos a seguir:

- Elegir la distribución del ruido deseado, (en este caso usamos  $m(t)$  basado en las observaciones 3.2).
- Calcular la función de distribución acumulada  $M(t)$  3.4.
- Normalizar  $M(t)$  3.5.
- Encontrar la función inversa de  $M_{norm}(t)$ .
- Evaluar  $t_i$  con  $i = 1, \dots, N$  puntos en  $M_{norm}^{-1}(t_i)$

El resultado es una serie de tiempo  $R(t)$  normalizada, de longitud  $N$  con distribución  $m(t)$ . Dado que podemos cambiar la SNR de la serie de tiempo simplemente con:  $R(t) * \sigma$ .

Para una curva normalizada ( $\mu = 1$ ) la SNR es igual a  $\sigma^{-1}$ . Con esto podemos elegir la SNR que se necesite.

Sin embargo, como mencionamos anteriormente, para representar adecuadamente las observaciones necesitamos una serie de tiempo de ruido simulado  $R(t)$  con un espectro no plano. Para obtener dicha serie de tiempo calculamos la FFT de  $R(t)$  y multiplicamos su espectro de frecuencias por una envolvente  $E(f)$ . El desglose detallado de los pasos a seguir es el siguiente:

- Calcular la FFT de  $R(t)$  la cual denotaremos como  $\mathcal{R}(f)$ .
- Generar de manera aleatoria los parámetros  $(C, a_1, b_1, a_2, b_2)$  distribuidos de manera uniforme dentro de los límites.
- Generar una envolvente evaluando los parámetros del paso anterior  $E((C, a_1, b_1, a_2, b_2), f)$ .
- Multiplicar  $\mathcal{R}(f) * E(f)$ .
- Calcular la FFT inversa para regresar al dominio del tiempo.
- Normalizar la nueva  $R(t)$ .
- Elegir  $\sigma$  para obtener la SNR deseada.

Esto resulta en una serie de tiempo, con  $SNR = \sigma^{-1}$ , con algunos ruidos sistemáticos aleatorios. En la figura 3.10 se muestra una comparación del espectro de frecuencias de la curva de luz de WASP-74b comparado con el ruido simulado, usando la envolvente  $E(f)$  calculada de su espectro.

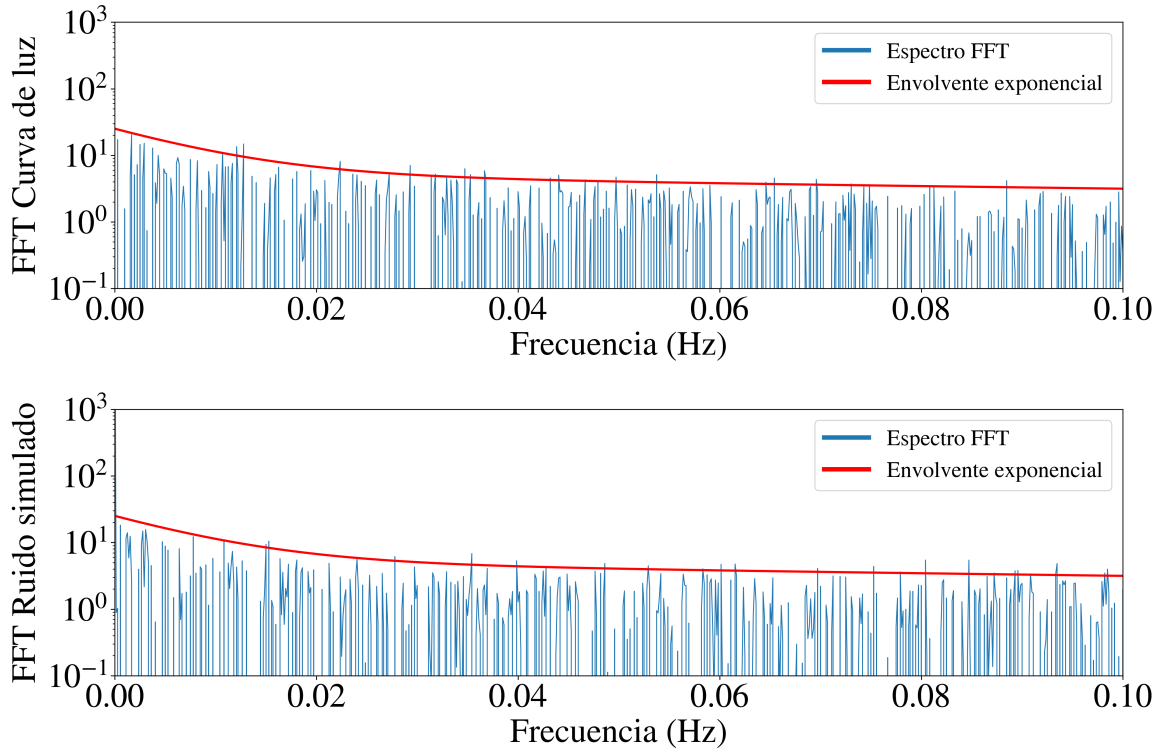


Figura 3.10: Arriba: se muestra la FFT de la curva de luz de WASP-74b. Abajo: el espectro de frecuencias de la serie de tiempo  $R(t)$  generada utilizando la metodología planteada en esta sección. Ambas gráficas son un acercamiento a la zona de bajas frecuencias para mejorar la visualización. La línea roja es la misma envolvente exponencial  $E(f)$  para ambos espectros.

Los valores para la amplitud ( $A_1, A_2$ ), las medias ( $\mu_1, \mu_2$ ) y las desviaciones ( $\sigma_1, \sigma_2$ ) de la función de distribución  $m(t)$  y la asíntota  $C$ , las amplitudes  $a_1$  y  $a_2$  y los factores  $b_1$  y  $b_2$  de la envolvente exponencial  $E(f)$  son elegidos de manera arbitraria dentro de un rango específico, el cual se obtuvo a partir de ajustar  $m(t)$  y  $E(f)$  a los datos observacionales.

Utilizando los resultados de 3.3, se creó una base de datos de ruidos simulados para asegurar reproducibilidad y posteriormente hacer estadística. Se variaron los parámetros ( $C, a_1, b_1, a_2, b_2$ ) de manera uniforme dentro de los intervalos:

$$\begin{aligned}
0.5 &\leq C \leq 50 \\
0 &\leq a_1 \leq 30 \\
0.02 &\leq b_1 \leq 2.5 \\
0 &\leq a_2 \leq 180 \\
0.002 &\leq b_2 \leq 0.06
\end{aligned} \tag{3.6}$$

estos intervalos se calcularon a partir de los espectros de los datos observacionales. Se generaron 100 curvas de ruido por cada valor de SNR. Los valores de SNR utilizados fueron 10, 20, 50, 80 y 100.

### III.4 Simulación de tránsitos de exoplanetas

Como se mencionó anteriormente (sección II.1), la curva de luz de un tránsito de exoplaneta depende de múltiples factores como el tamaño de la estrella, el tamaño del planeta, la velocidad orbital del planeta, el ángulo con respecto a la línea de visión etc.

Para la simulación de curvas de luz en este trabajo, se utilizó la paquetería BATMAN (del inglés *B*asic *T*ransit *M*odel *c*Alculation) desarrollada en *Python* (Kreidberg, 2015). Este algoritmo utiliza el modelo analítico propuesto por (Mandel and Agol, 2002)

#### III.4.1 Descripción y uso de BATMAN

BATMAN es un paquete para *Python*, con el cual se modelan tránsitos de exoplanetas y sus curvas de luz. El código usa extensiones en C para aumentar la velocidad de cálculo y está paralelizado con OpenMP (Kreidberg, 2015). BATMAN puede calcular curvas de luz para cualquier ley de oscurecimiento de limbo radial, usando un nuevo algoritmo de integración que no puede ser calculado analíticamente.

Para utilizar BATMAN, se requiere ingresar los parámetros físicos de los cuales depende el tránsito, tales como: el periodo orbital ( $P$ ), el radio del planeta ( $R_p$ ) y el semieje mayor ( $a$ ), ambos en unidades del radio de la estrella anfitriona ( $R_*$ ); la excentricidad de la órbita ( $\epsilon$ ), la inclinación orbital ( $i$ ) y la longitud del periastro ( $\bar{\omega}$ ) ambos en grados. Además se puede seleccionar la ley de oscurecimiento radial (véase II.1.2). En la figura

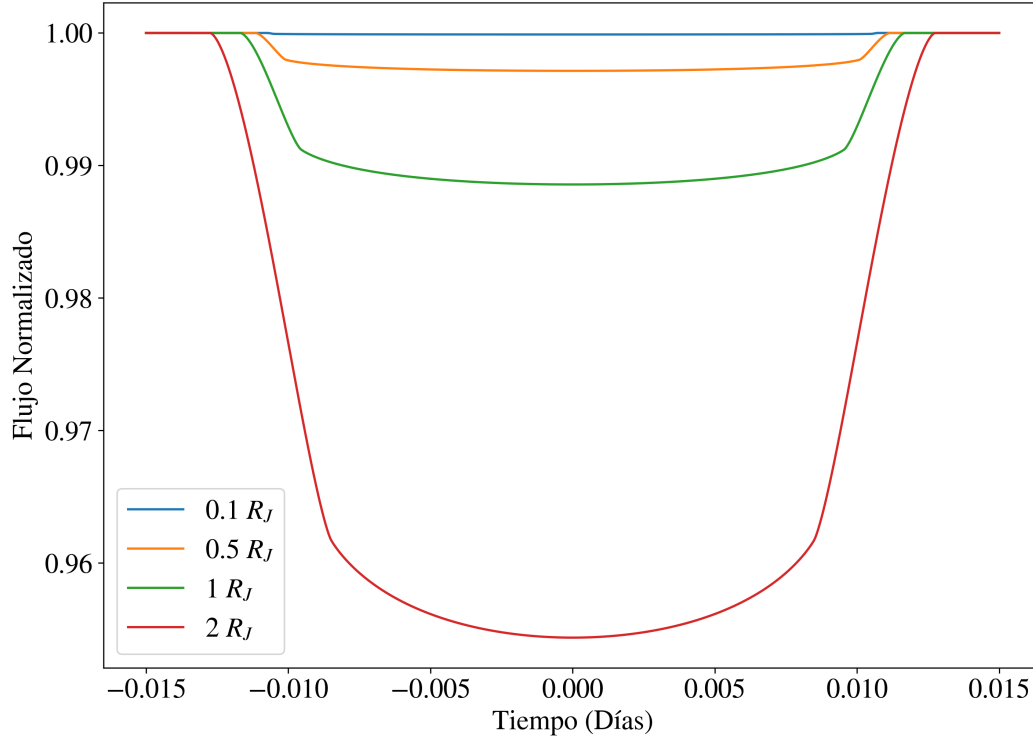


Figura 3.11: Curvas de luz de tránsitos de exoplanetas con  $P = 1$  día,  $i = 90^\circ$ ,  $a = 15R_*$ ,  $\epsilon = 0$ ,  $\bar{\omega} = 90$  y un modelo de oscurecimiento lineal. Las distintas curvas representan la diferencia en el cambio del flujo  $\Delta F$  para diferentes valores de  $R_p$ .

3.11 se muestra un ejemplo de curvas de luz simuladas con BATMAN.

Utilizando BATMAN se creó una base de datos de tránsitos simulados, con un total de 625 curvas diferentes, utilizando 25 diferentes valores de las observables  $\Delta F$  y  $t_c$ . Los parámetros orbitales fueron elegidos de tal manera que obtuvimos valores entre 0.5 %-5 % para  $\Delta F$  y entre 15-60 minutos para  $t_c$ .

### III.5 Reducción de ruido en curvas de luz

Como se mencionó con anterioridad, para la detección de un tránsito exoplanetario se necesita poder medir variaciones de flujo  $\Delta F \sim 1 - 2\%$ . Lo cual en las curvas de luz de

alta cadencia es difícil debido a la cantidad de ruido. En las subsecciones siguientes, se muestran los métodos para mejorar la SNR y poder identificar la presencia de un tránsito en una curva de luz.

### III.5.1 Reducción del ruido usando promedio móvil

En la sección II.2 se explicó la manera convencional en la que los proyectos enfocados a la detección de exoplanetas realizan sus observaciones. Todos ellos utilizan imágenes de mediana-larga exposición, para así aumentar la SNR. Se puede intentar simular una curva de luz de larga exposición desde una de alta cadencia, simplemente integrando la contribución de  $n$  muestras de la curva de luz.

Por ejemplo, las curvas de luz de este trabajo fueron tomadas a una cadencia de 20 fps, es decir, 0.05 segundos de exposición, durante aproximadamente 2 horas, lo cual resulta en una curva  $F(t)$  de  $N = 144,000$  muestras. Entonces, promediando el flujo de  $n = 1200$  muestras de la curva de luz  $F(t)$ , se puede intentar simular una muestra obtenida a partir de una imagen con 60 segundos de exposición  $F_{prom}(t)$ . Es decir, cada  $n = 1200$  muestras de la curva de luz original, resultan en  $N_{prom} = 1$  muestras en la curva promediada:

$$F_{prom}(t) = \frac{\sum_{i=m}^{m+n} F(t_i)}{n} \quad m = 0, n, 2n, \dots, N - n; \quad (3.7)$$

Esto resulta en una curva con longitud  $N_{prom} = N/n$ . Esta misma metodología aplicado a las imágenes, es el método Shift & Add de (Klückers et al., 1996). Sin embargo, el ruido electrónico resultante de la lectura de cada imagen es imposible de remover con esta técnica, por lo tanto, una curva de luz obtenida de imágenes con 60 segundos de exposición es siempre menos ruidosa que una obtenida promediando 1200 imágenes de 0.05 segundos de exposición.

Para evitar perder resolución temporal, se utiliza la técnica promedio móvil (Borucki et al., 2009). Esta técnica consiste en igualmente promediar la contribución de  $n$  muestras, pero desplazando el promediado 1 muestra, en vez de  $n$  como en el ejemplo anterior, es decir:

$$F_{mov}(t) = \frac{\sum_{i=m}^{m+n} F(t_i)}{n} \quad m = 0, 1, 2, \dots, N - n; \quad (3.8)$$

Lo cual resulta en una curva de luz  $F_{mov}(t)$  suavizada de longitud  $N_{mov} = N - n$ . Otra opción análoga a este método, es usar la mediana, en lugar del promedio, esta tiene un comportamiento mas suave cuando se tienen valores atípicos. Un ejemplo del efecto de estos métodos en la curva de de luz de WASP-74 (figura 3.5) se muestra en la figura 3.12.

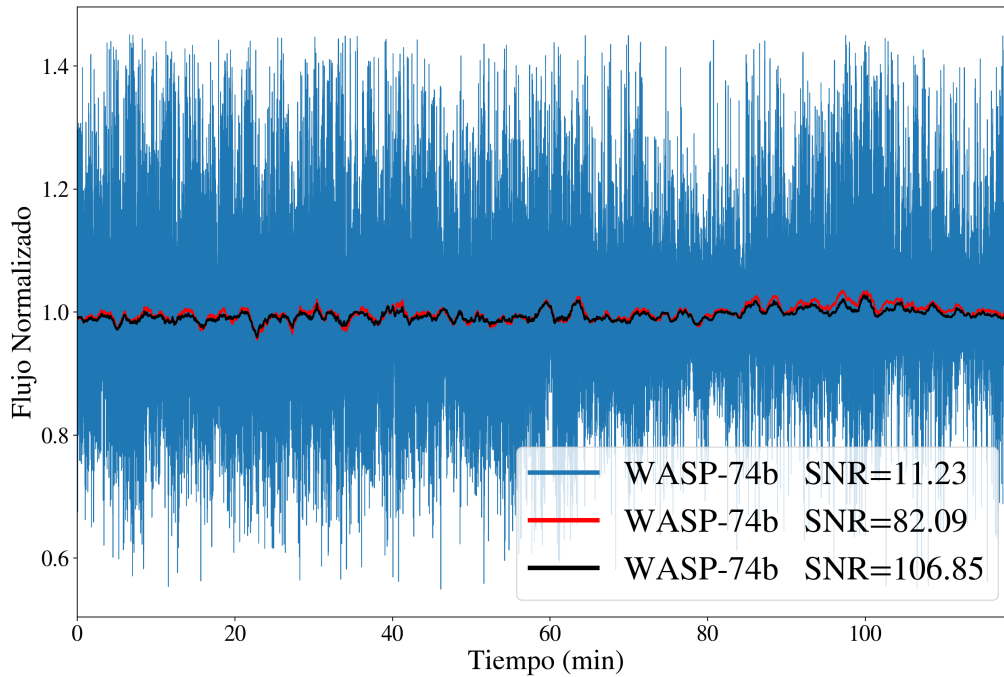


Figura 3.12: Curva de luz de WASP-74 tomada el día 28 de mayo del 2017. En azul se muestra la curva de luz original obtenida mediante fotometría de apertura. En rojo la curva resultado del promedio móvil, y en negro la curva resultado de tomar la mediana en lugar del promedio. En el recuadro se muestra la SNR calculada para cada curva de utilizando (2.11).

Esta metodología es útil para el suavizado de curvas en general. Sin embargo, de la misma forma que en el promediado fijo, no podemos asumir que la curva  $F_{mov}(t)$  con



$n = 1200$  es equivalente a una obtenida utilizando un tiempo de exposición de 60 segundos, ya que el ruido electrónico y de lectura del detector tienen una contribución significativa en las curvas de alta cadencia, lo cual puede añadir ciertos sistemáticos.

### III.5.2 Identificación de firma de tránsito usando filtro pasa bajas

Un tránsito inmerso en una curva de luz de alta cadencia, puede interpretarse como una variación de baja frecuencia en el brillo de la estrella. Partiendo de esta suposición, la firma del tránsito, debe poder distinguirse en el espectro de frecuencias, particularmente en la zona de bajas frecuencias ( $\sim 1 \times 10^{-4} Hz$ ).

En una curva de luz de alta cadencia, se pueden observar a simple vista variaciones de alta frecuencia, las cuales ocultan el tránsito. Estas contribuciones de alta frecuencia pueden ser atenuadas utilizando un filtro pasa bajas, específicamente un filtro pasa-bajas (Thompson, 1983). La Figura 3.13 muestra un ejemplo del filtro pasa-bajas utilizado.

Para aplicar el filtro  $\mathcal{H}(f)$  a  $F(t)$ , primero se calcula su FFT para obtener:  $\mathcal{F}(f)$ . Después multiplicamos el filtro, por el espectro de frecuencias para obtener el espectro filtrado  $\mathcal{F}_{fil}(f)$ :

$$\mathcal{F}_{fil}(f) = \mathcal{H}(f) \times \mathcal{F}(f) \quad (3.9)$$

finalmente regresamos al dominio del tiempo con la FFT inversa. Este proceso es análogo a utilizar una función  $Sinc(t)$  como función ventana y usar la convolución en el dominio del tiempo. Este proceso es conocido como *Windowed Fast Fourier Transform* (WFFT) (Owen, 2007). Sin embargo, el proceso descrito anteriormente es computacionalmente más eficiente y obtenemos los mismos resultados, un ejemplo de la aplicación de este filtrado se muestra en la figura 3.14.

Este filtrado es llamado ideal por su simplicidad, sin embargo, no hay necesidad de utilizar otro tipo de funciones para  $\mathcal{H}(f)$ , ya que las altas frecuencias ( $> 10 Hz$ ) no aportan información de la firma del tránsito exoplanetario (Mahu et al., 2017), debido al ser un evento de larga duración comparado a la frecuencia de muestreo (20 imágenes por segundo).

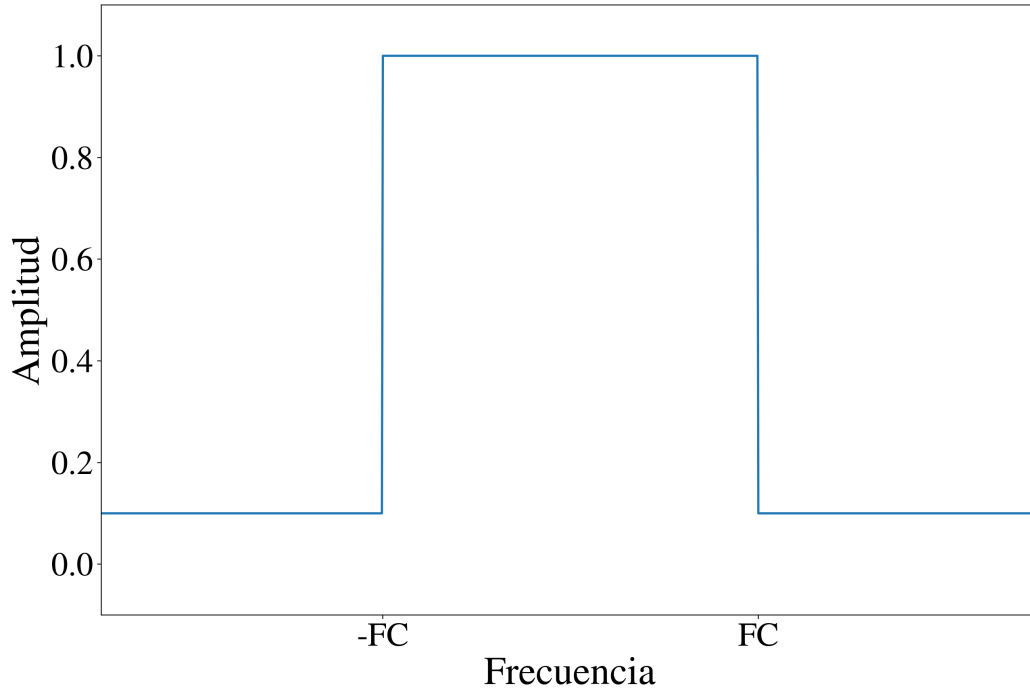


Figura 3.13: Filtro pasa-bajas  $\mathcal{H}(f)$  utilizado para remover altas frecuencias.  $FC$  representa la frecuencia de corte la cual puede variar dependiendo la curva de luz.

### III.5.3 Identificación de firma de tránsito usando PCA

Para toda matriz  $A$  con dimensiones  $m \times n$ , su descomposición en valores singulares ó SVD (del inglés *Singular Value Decomposition*) (Golub and Reinsch, 1971) es:

$$A = U\Sigma V^T; \quad (3.10)$$

donde:

- $U$  es llamado *vector singular* y es ortogonal ( $UU^T = I$ ),
- $\Sigma$  está compuesta por los valores singulares de  $U$  y es diagonal no negativa,
- $V$  es también un *vector singular* y es ortogonal ( $VV^T = I$ ).

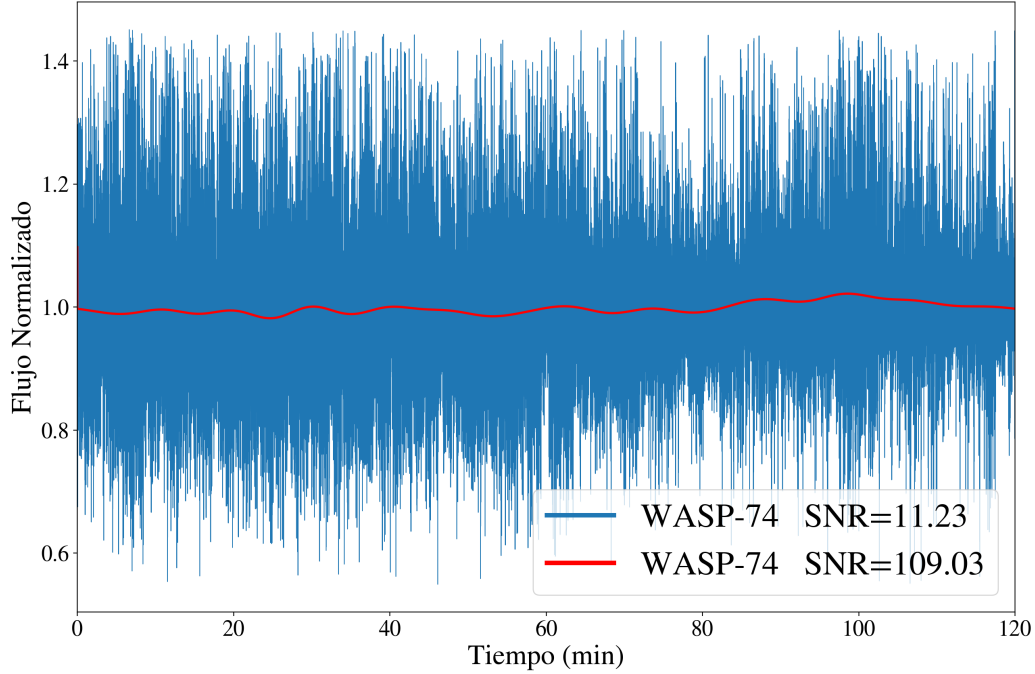


Figura 3.14: Curva de luz de WASP-74 tomada el día 28 de mayo del 2017. En azul se muestra la curva de luz original obtenida mediante fotometría de apertura. En rojo la curva resultado del filtrado de Fourier, utilizando una frecuencia de corte  $FC = 2 \times 10^{-3}$ .

Si multiplicamos la ecuación 3.10 por su transpuesta tenemos:

$$AA^T = U\Sigma V^T V \Sigma U^T = U\Sigma^2 U^T; \quad (3.11)$$

el resultado es el conocido teorema de diagonalización para matrices simétricas que dice: *Sea  $S$  una matriz cuadrada de  $M \times M$  con  $M$  eigenvectores linealmente independientes, entonces existe una descomposición simétrica tal que:*

$$S = Q\Lambda Q^T;$$

donde las columnas de  $Q$  son los eigenvectores normalizados de  $S$ , y  $\Lambda$  es la matriz diagonal con los eigenvalores de  $S$ .

El significado matemático de la SVD, depende de la forma de  $A$ . Normalmente  $A$  con dimensiones  $m \times n$  está compuesta por los datos, donde  $m$  es el número de muestras y  $n$  el número de variables. Los *vectores singulares*  $U$ , son también conocidos como *ejes principales* y la proyección de los datos  $A$  sobre estos ejes principales son llamados *componentes principales* ( $P_{comp}$ ), es decir ( $P_{comp} = AU$ ), de ahí el nombre de análisis de componentes principales ó PCA (del inglés *Principal Component Analysis*).

Esta descomposición es utilizada para la extracción de características y disminuir la dimensionalidad de los datos que se estén analizando (McGurk et al., 2010); (Medeiros et al., 2018). En este trabajo se utilizan las componentes principales para limpiar la curva de luz del ruido sistemático (Shin et al., 1999); (Bailey, 2012).

Existen metodologías para reducir el ruido en series de tiempo utilizando análisis de componentes principales (PCA). Sin embargo, el PCA puede emplearse de manera más conveniente, si se conocen las características de la señal que buscamos inmersa en el ruido, en este caso, ingresos o egresos de tránsitos de exoplanetas.

Primero se necesita construir una matriz  $X$  con muestras  $\{x_i | i = 1, \dots, N\}$ . Cada una de estas muestras  $x_i$ , es una curva de luz de un tránsito simulado con BATMAN, con distintos periodos ( $P$ ) y radios de planeta ( $R_p$ ):

$$X = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{pmatrix} = \begin{pmatrix} v_1^1 & v_2^1 & \cdots & v_n^1 \\ v_1^2 & v_2^2 & \cdots & v_n^2 \\ \vdots & \vdots & \ddots & \vdots \\ v_1^N & v_2^N & \cdots & v_n^N \end{pmatrix} \quad (3.12)$$

donde  $n$  es el número de puntos en la curva simulada  $\{v_i^N | i = 1, \dots, n\}$ . Por cual entonces,  $X$  tiene dimensiones  $N \times n$ , y su SVD está dada por:

$$X = U\Sigma V^T; \quad (3.13)$$

Para calcular la matriz de eigenvalores  $V^T$ , se utilizó la función `TruncatedSVD` de la paquetería *Scikit Learn* en *Python*. Uno puede cuantificar cuales de las componentes principales son más representativas de las muestras en la matriz  $X$ . Como su nombre lo indica, esta función nos regresa solamente las primeras  $k$  componentes más significativas, donde  $k$  s definido por el usuario. Esta función nos proporciona la matriz  $V^T$  truncada,

de dimensiones  $k \times N$ .

Después se calcula la contribución de las  $k$  componentes sobre una curva de luz  $F_t$  o un conjunto de  $l$  curvas de luz:

$$C = V^T \times F_t; \quad (3.14)$$

donde  $C$  tiene dimensiones  $k \times l$ , si utilizamos una sola curva de luz  $l = 1$ . Finalmente para obtener la curva de luz  $F'_t$  reconstruida en base a las  $k$  componentes principales:

$$F'_t = C^T \times V^T; \quad (3.15)$$

Este proceso utiliza las componentes principales de los tránsitos simulados para reconstruir la curva de luz, resaltando las componentes principales. Esto no es análogo a una reducción de ruido. Si la curva  $F_t$  que se va analizar no contiene las características que se están buscando (las características de  $X$ ), el resultado es una curva  $F'_t$  que no posee ningún patrón específico. Esto no se convierte en un problema, dado que solo estamos interesados a las curvas que sí presenten las características de un tránsito.

### *III.6 Criterio de validación para candidatos a tránsitos de exoplanetas*

Después de aplicar las metodologías para mejorar la SNR, se necesita un proceso para validar la presencia de un posible tránsito de exoplaneta. Para esto, se ajustó un modelo trapezoidal a los datos filtrados, y basados en la calidad del ajuste, se etiquetan las curvas de luz como posibles candidatos.

#### *III.6.1 Modelo del trapezoide*

Las curvas de luz de alta candencia son mucho más tardadas de procesar por el número de mediciones que se realizan. Teniendo en cuenta la eficiencia del algoritmo, se utilizó el modelo más simple para un tránsito de exoplaneta, un trapezoide. Este modelo es muy utilizado en el procesamiento de señales producidas por tránsitos de exoplanetas, por ejemplo los trabajos (Alapini Odunlade, 2010), (Hippke and Heller, 2019), (Kipping and Sandford, 2016) y (Morton, 2012).

Con este modelo reducimos la cantidad de parámetros a tres: la profundidad del

tránsito  $\Delta F$ ,  $D$  el punto de comienzo del evento (ingreso o egreso del planeta al disco solar.) y la duración del evento  $t_c$ .

Como se explicó en la sección II.1, a partir de curvas de luz de tránsitos solo podemos conocer el periodo orbital  $P$  y la profundidad del tránsito  $\Delta F$ , la cual está en función del radio del planeta  $R_p$  y el radio de la estrella  $R_*$ . Por lo que el modelo del trapecoide es suficiente para definir un posible candidato a tránsito.

### III.6.2 Coeficiente de correlación de Pearson

Después de ajustar el modelo del trapecoide a las curvas de luz, se utilizó el coeficiente de correlación de Pearson, para evaluar la calidad del ajuste.

$$C = \frac{\sigma_{12}}{\sigma_1\sigma_2}; \quad (3.16)$$

donde  $\sigma_{12}$  es la covarianza entre la curva de luz y el modelo ajustado, definida como:

$$\sigma_{xy} = \sum_i^n \frac{1}{n} (x_i - \bar{x})(y_i - \bar{y}); \quad (3.17)$$

donde  $\sigma_1$  y  $\sigma_2$  son la desviación estándar de la curva de luz y el modelo respectivamente. A partir de simulaciones, se calculó el limite inferior para el coeficiente de correlación, para evaluar si la estrella y su curva de luz se cataloga como objeto de interés, para elaborar un análisis más detallado (véase IV.1.2)

## CAPÍTULO IV: ANÁLISIS DE RESULTADOS

### *IV.1 Simulaciones*

#### *IV.1.1 Determinación de candidatos a tránsitos*

Para determinar la eficiencia de la metodología en la identificación de candidatos a tránsitos de exoplanetas, se utilizó el coeficiente de correlación de Pearson como medida de la calidad del ajuste. También evaluamos la distorsión de la curva de luz, debido a los diferentes métodos de filtrado, para esto calculamos la diferencia porcentual entre la duración del ingreso-egreso del planeta ( $t_c$ ) de la simulación y los obtenidos mediante el ajuste.

Utilizando la base de datos de tránsitos simulados (véase III.3.1) y los ruidos simulados (véase III.2.3) se calcularon los coeficientes de correlación de 625 casos únicos ( $25 \Delta F$  y  $25 t_c$ ).

Presentaremos los resultados por método de filtrado, y para los casos extremos de ruido, es decir  $SNR = 10$  y  $SNR = 100$ . Estos valores de SNR's se calcularon antes de cualquier proceso de filtrado.

##### *IV.1.1.1 Resultados del filtro pasa bajas*

La figura 4.1 muestra la correlación, entre el modelo y la curva filtrada con el filtro pasa bajas, para curvas con  $SNR = 10$  y  $SNR = 100$  respectivamente. El color nos indica la media de los coeficientes de correlación ( $C$ ), de los 100 ruidos con diferentes parámetros (véase III.2.3).

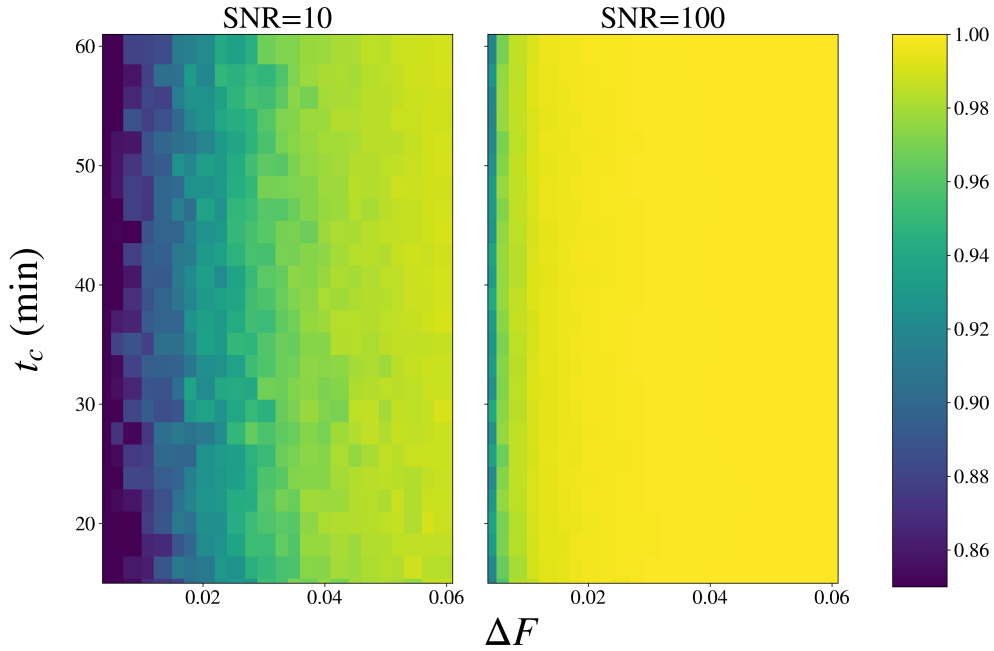


Figura 4.1: Matriz de correlaciones de los parámetros  $\Delta F$  y  $t_c$  entre las simulaciones y los calculados después de aplicar el modelo del trapecioide a los datos filtrados utilizando Fourier. Estos resultados se obtuvieron a partir de curvas simuladas con una  $SNR = 10$  y  $SNR = 100$ .

Estos resultados ayudan a saber que tipos de sistemas planetarios, podrían ser detectados utilizando esta metodología. También nos ayuda a definir la cota inferior para el coeficiente de correlación entre el modelo y la curva filtrada, para que esta, sea considerada como candidato a tránsito. Para el método de filtrado con Fourier, la cota inferior para el coeficiente de correlación es de  $C \geq 0.85$ .

Esta cota inferior se calculó a partir de un estudio de falsos positivos, donde se aplicó la metodología planteada en este trabajo, a curvas con solo ruido. De estos resultados, tomamos el coeficiente de correlación máximo  $C_{max}$  y calculamos la cota inferior como:

$$C_{cota} = \frac{1 + C_{max}}{2} \quad (4.1)$$

Si analizamos la figura 4.1, vemos que para las curvas con  $SNR = 100$ , todo el intervalo de valores de  $\Delta F$  y  $t_c$  (dentro de los límites, véase IV.2) tiene un valor de  $C$  es mayor



que la cota inferior. Esto significaría, que en principio seríamos capaces de detectarlos. Por otra parte, en la figura 4.1, vemos que para curvas con  $SNR = 10$  los tránsitos con  $\Delta F \lesssim 0.1$  no serían detectables.

La figura 4.2 muestra la diferencia porcentual entre el valor calculado para la duración del evento ( $t_c^{Fit}$ ) utilizando el modelo del trapecoide y el valor inicial de la simulación ( $t_c^{Real}$ ). Se ilustran también los resultados para una  $SNR = 10$  y los de  $SNR = 100$ , utilizando la misma escala para facilitar la comparación. Aquí podemos observar la mejora en el cálculo de la duración del evento, debido a someter la curva de luz a un filtro pasa bajas.

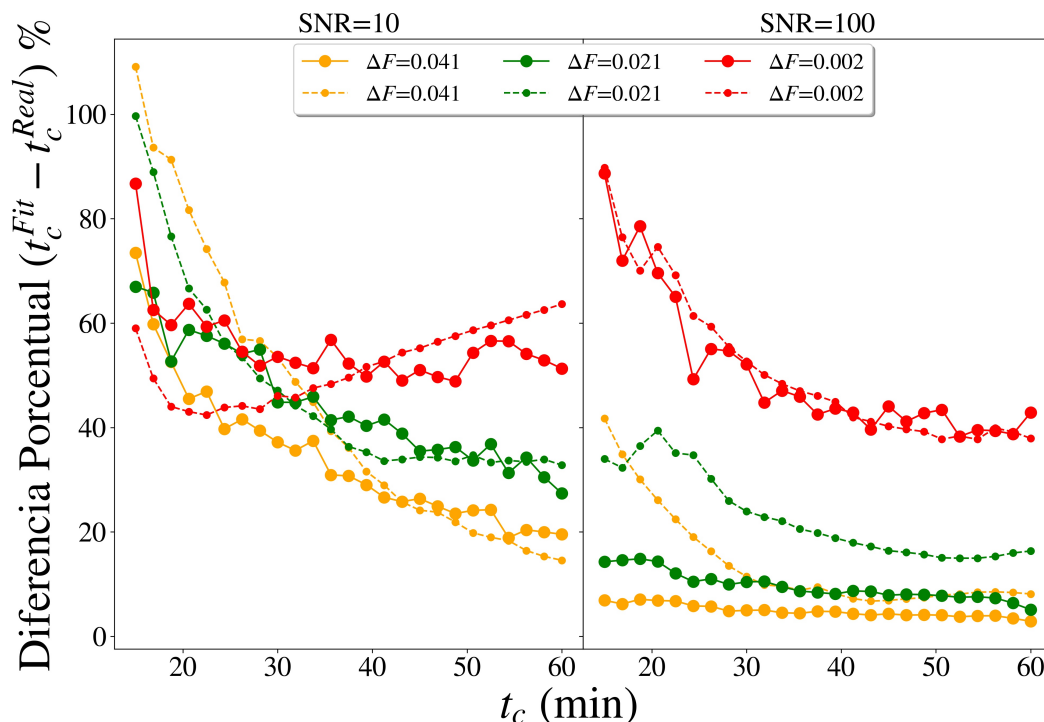


Figura 4.2: Las líneas sólidas representan la diferencia porcentual entre  $t_c^{Fit}$  y  $t_c^{Real}$  después de aplicar el modelo del trapezoide a los datos filtrados utilizando el filtro pasa bajas. Las líneas punteadas representan la misma cantidad, pero obtenida aplicando el modelo del trapezoide a las curvas simuladas sin filtrar. El color representa diferentes valores de  $\Delta F$ , se utilizaron 4%, 2% y 0.2% representados en amarillo, verde y rojo respectivamente. Estos resultados se obtuvieron a partir de curvas simuladas a las que se les insertó un ruido con una  $SNR = 10$  (izquierda) y  $SNR = 100$  (derecha).

Podemos observar que para tránsitos con una profundidad  $\Delta F > 1\%$  y los  $t_c$  más cortos (15-40 min) este método es más efectivo, fuera de estos rangos no se aprecia una diferencia significativa entre los resultados obtenidos sin aplicar filtrado y aplicándolo. En el caso de  $SNR = 10$  se aprecia como para el caso de  $\Delta F = 0.2\%$  en algunos casos, el ajuste sin filtrar dio mejores resultados, esto se debe a la deformación en los bordes de la curva debido a el filtrado de Fourier.

## IV.1.1.2 Resultados del promedio móvil

De manera análoga, la figura 4.3 muestran la correlación, entre el modelo y la curva filtrada con el promedio móvil, para curvas con  $SNR = 10$  y  $SNR = 100$  respectivamente. El color nos indica la media de los coeficientes de correlación ( $C$ ), de los 100 ruidos con diferentes parámetros (véase III.2.3).

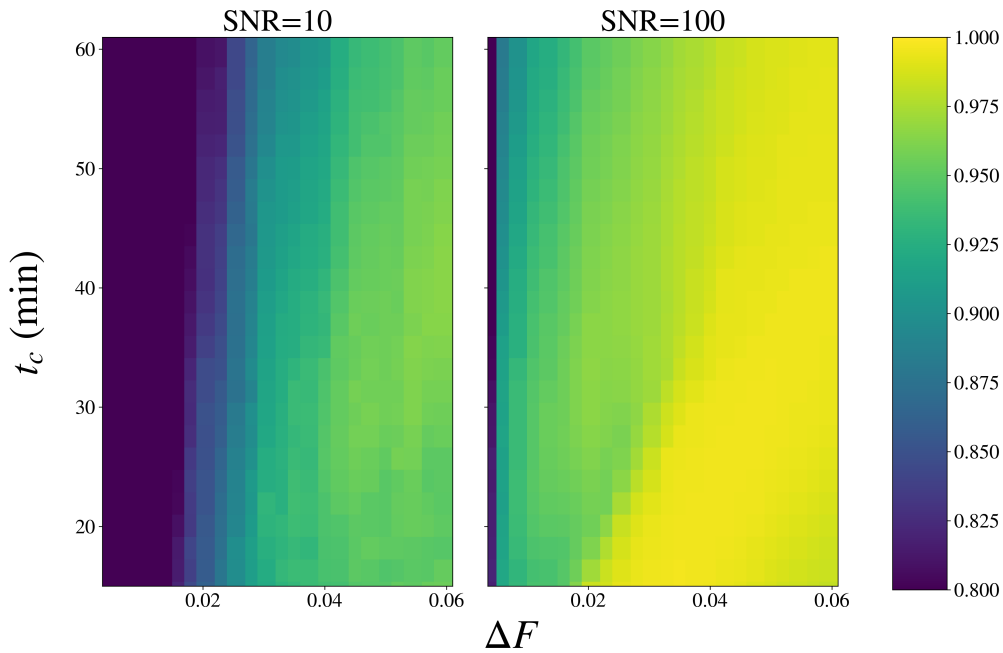


Figura 4.3: Matriz de correlaciones de los parámetros  $\Delta F$  y  $t_c$  entre las simulaciones y los calculados después de aplicar el modelo del trapecoide a los datos filtrados utilizando promedio móvil. Estos resultados se obtuvieron a partir de curvas con una  $SNR = 10$  y  $SNR = 100$ .

Para el método de promedio móvil, la cota inferior para determinar una curva como candidata a tránsito es de  $C \geq 0.84$ . Si analizamos la figura 4.3, vemos que para este método de filtrado, los valores de  $C$  están en función de  $\Delta F$  y más notablemente de  $t_c$ .

La figura 4.4 muestran la diferencia porcentual entre  $t_c^{Fit}$  y  $t_c^{Real}$ , que representan la duración de  $t_c$  calculada mediante el ajuste y el valor real de la simulación respectivamente. Podemos observar que los resultados son consistentes con la matriz de correlaciones.

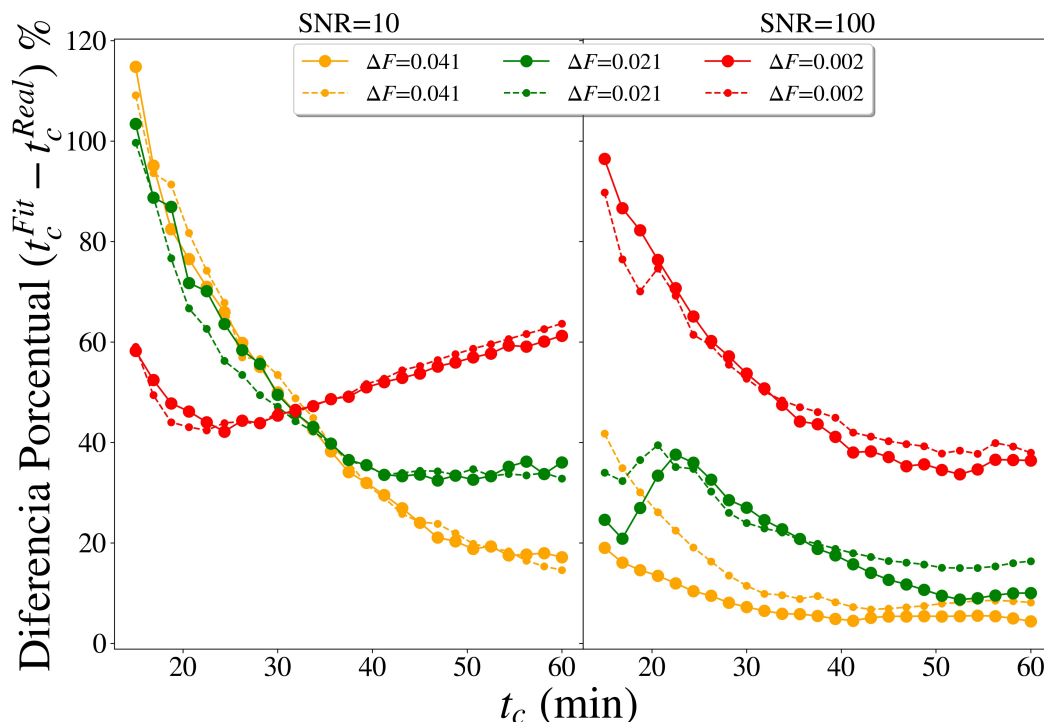


Figura 4.4: Las líneas sólidas representan la diferencia porcentual entre  $t_c^{Fit}$  y  $t_c^{Real}$  después de aplicar el modelo del trapecoide a los datos filtrados utilizando promedio móvil. Las líneas punteadas representan la misma cantidad, pero obtenida aplicando el modelo del trapecoide a las curvas simuladas sin filtrar. El color representa diferentes valores de  $\Delta F$ , se utilizaron 4%, 2% y 0.2% representados en amarillo, verde y rojo respectivamente. Estos resultados se obtuvieron a partir de curvas simuladas a las que se les insertó un ruido con una  $SNR = 10$ .

Podemos observar que este método es el menos efectivo, si queremos obtener una medida precisa de  $t_c$ . No se aprecia una diferencia significativa entre los resultados para las mismas curvas con y sin filtrar. Esto se debe a lo que se mencionaba con anterioridad, este método altera la forma de la señal inmersa en el ruido. Al promediar en el tiempo, la pendiente del ingreso-engreso del planeta en la curva de luz cambia (véase 2.2). Esto provoca que aún suavizando la curva el valor real difiere ligeramente del calculado. Para minimizar este efecto, se elige una ventana de promediado pequeña, para estos resultados se utilizó una ventana de promediado de  $n = 2000$  muestras.

Además esta deformación se vuelve más significativa de manera proporcional a  $t_c$ . Este método es el menos eficiente computacionalmente hablando.

#### IV.1.1.3 Resultados del filtrado con PCA

De manera análoga, la figura ?? muestran la correlación, entre el modelo y la curva filtrada con el método de análisis de componentes principales (PCA), para curvas con  $SNR = 10$  y  $SNR = 100$  respectivamente. El color nos indica la media de los coeficientes de correlación ( $C$ ), de los 100 ruidos con diferentes parámetros (véase III.2.3).

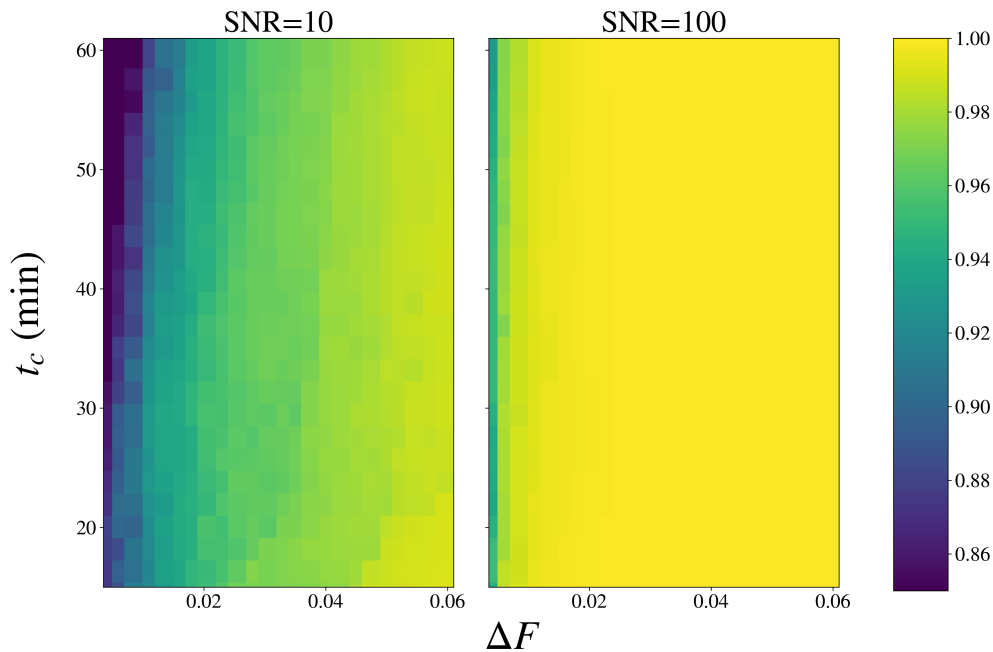


Figura 4.5: Matriz de correlaciones de los parámetros  $\Delta F$  y  $t_c$  entre las simulaciones y los calculados después de aplicar el modelo del trapecioide a los datos filtrados utilizando PCA. Estos resultados se obtuvieron a partir de curvas con una  $SNR = 10$  y  $SNR = 100$ .

El filtrado usando PCA método arroja los coeficientes de correlación más altos, sin embargo lo mismo ocurrió en la prueba de falsos positivos. Para este método, la cota inferior para determinar una curva como candidata a tránsito es de  $C \geq 0.94$ .

Si analizamos la figura 4.5, vemos que al igual que para el método de Fourier, no se

ve una correlación significativa entre  $C$  y  $t_c$ .

La figura 4.6 muestran la diferencia porcentual entre  $t_c^{Fit}$  y  $t_c^{Real}$ , que representan la duración de  $t_c$  calculada mediante el ajuste y el valor real de la simulación respectivamente.

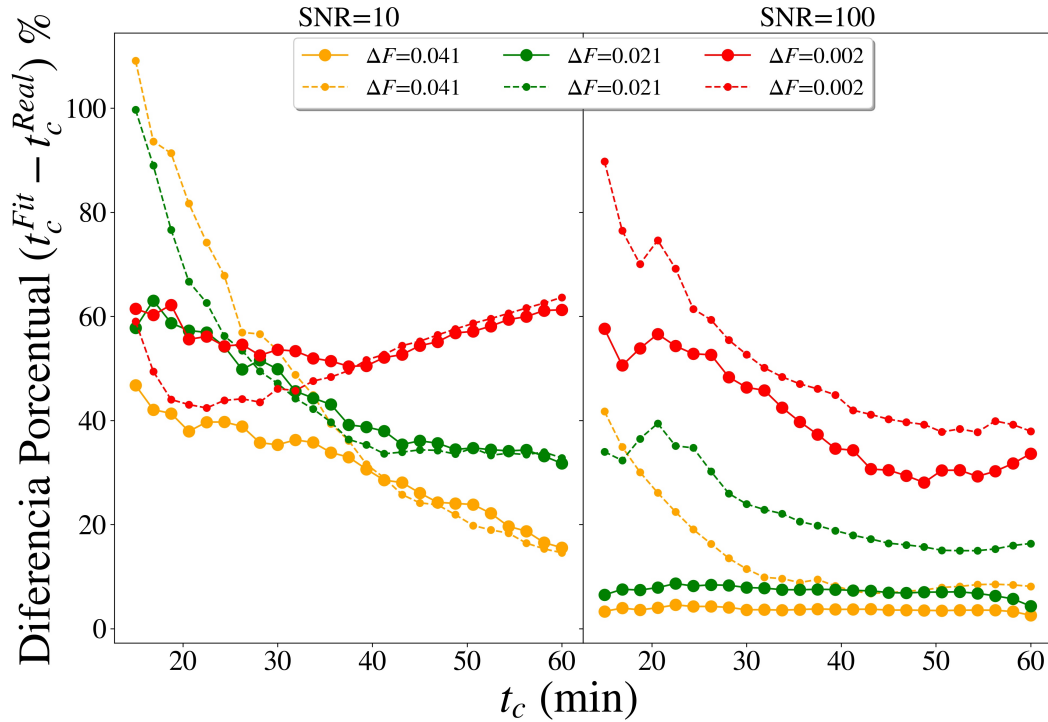


Figura 4.6: Las líneas sólidas representan la diferencia porcentual entre  $t_c^{Fit}$  y  $t_c^{Real}$  después de aplicar el modelo del trapecoide a los datos filtrados utilizando PCA. Las líneas punteadas representan la misma cantidad, pero obtenida aplicando el modelo del trapecoide a las curvas simuladas sin filtrar. El color representa diferentes valores de  $\Delta F$ , se utilizaron 4%, 2% y 0.2% representados en amarillo, verde y rojo respectivamente. Estos resultados se obtuvieron a partir de curvas simuladas a las que se les insertó un ruido con una  $SNR = 10$  (izquierda) y  $SNR = 100$  (derecha).

Podemos observar que este método es el más efectivo, si queremos obtener una medida precisa de  $t_c$ . De igual manera que el filtro pasa bajas, este funciona mejor para pendientes más grandes, es decir tiempos de ingreso-egreso ( $t_c$ ) cortos ( $\lesssim 30$  min). Se aprecia una mejora debido al filtrado para todos los valores de  $t_c$  del al menos 15%. También cabe destacar que este método de filtrado es el más eficiente, esto desde el punto de vista

computacional.

## *IV.2 Datos observacionales*

Se presentan los resultados de las metodologías para la mejora de la SNR en las curvas de luz reales (véase III.1).

### *IV.2.1 Mejora de la señal a ruido en datos observacionales*

Se le aplico la metodología propuesta en este trabajo a todas las curvas de luz (véase tabla 3.2, sin embargo para mejorar la visualización, presentamos solamente los resultados del filtrado de ruido en la curva de luz de alta cadencia de HAT-P-37b, utilizando las 3 metodologías descritas en la sección anterior: el filtro de frecuencias de Fourier, promedio móvil y análisis de componentes principales (PCA). La curva de luz de HAT-P-37b fue obtenida mediante fotometría de apertura utilizando APPHi.

#### *IV.2.1.1 Resultados del filtro pasa bajas*

En la figura 4.8 podemos apreciar el resultado del filtro pasa bajas de Fourier, aplicado a la curva de luz de HAT-P-37b. A pesar de ser una curva con mucho ruido, se puede apreciar a simple vista la forma del tránsito. Se utilizó una frecuencia de corte  $FC$  de  $2 \times 10^{-3}$ . También se muestra el modelo teórico del tránsito en negro, simulado usando los parámetros presentados en (Bakos et al., 2012) y *BATMAN* (véase III.3.1). La similitud entre el modelo y la curva filtrada es alta, el coeficiente de correlación es de  $C = 0.982$ .

#### *IV.2.1.2 Resultados del promedio móvil*

En la figura 4.9 podemos apreciar el resultado del promedio móvil, aplicado a la curva de luz de HAT-P-37b. Para este resultado, se utilizó una ventana de 3000 frames. De igual manera se muestra en negro el modelo teórico. La similitud entre el modelo y la curva filtrada es alta, el coeficiente de correlación es de  $C = 0.962$ .

Como se observa en la figura 4.9, este método altera la pendiente de la curva, donde el planeta está saliendo de la superficie solar. Esta deformación es proporcional a la longitud

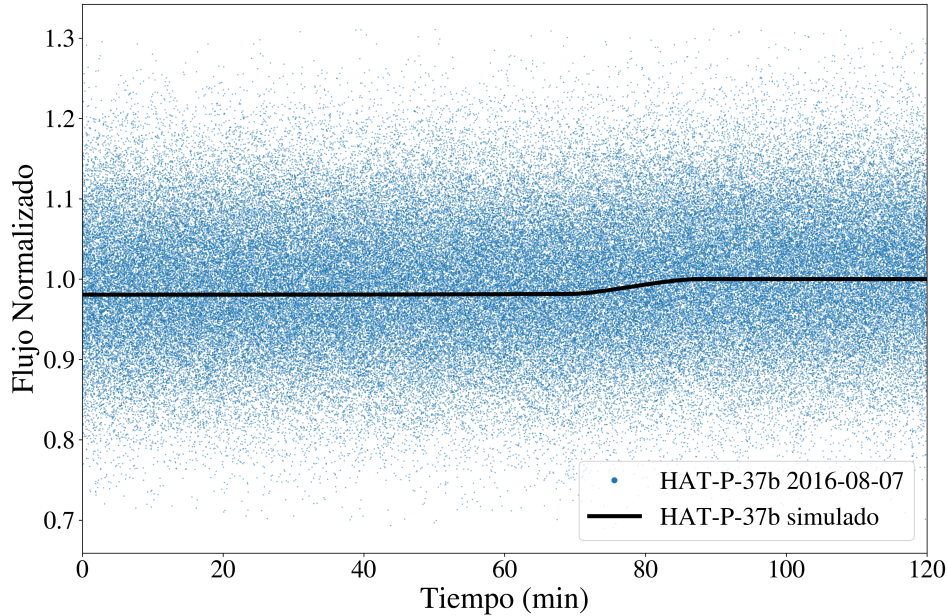


Figura 4.7: Curva de luz de HAT-P-37b tomada a 20 fps, durante la segunda mitad del tránsito. La curva de luz se obtuvo mediante fotometría de apertura con una  $SNR = 11.73$ . La línea negra representa el modelo teórico, utilizando los parámetros presentados en (Bakos et al., 2012).

de la ventana de promediado, por lo que ventanas pequeñas (1000-3000 frames) pueden dar mejores coeficientes de correlación.

#### IV.2.1.3 Resultados de PCA

En la figura 4.10 podemos apreciar el resultado del filtrado utilizando el análisis de componentes principales (PCA), aplicado a la curva de luz de HAT-P-37b. Se calcularon las componentes principales de un conjunto de 25 tránsitos simulados con diferentes  $\Delta F$  y  $t_c$ . Debido a que las componentes principales son calculadas utilizando una variedad de curvas simuladas, la curva que resulta después del filtrado posee la forma de tránsito, sin embargo, no es ideal para ser utilizada para el cálculo de observables.

Esta metodología funciona para la visualización, y como método de comparación entre los modelos teóricos y las curvas de luz. Si la curva de luz, contiene componentes prin-



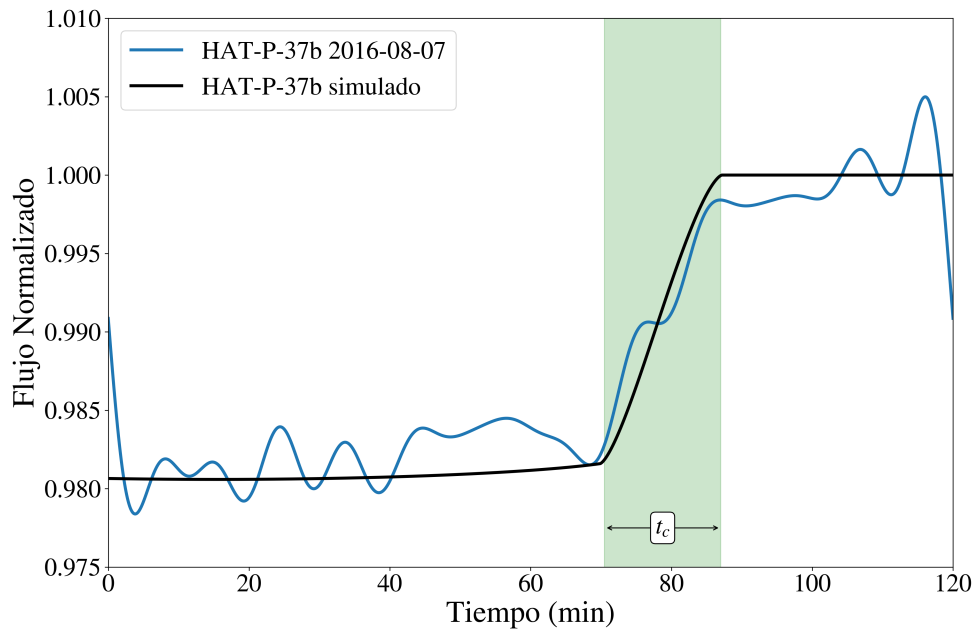


Figura 4.8: Azul: curva de luz de HAT-P-37b después del filtrado de altas frecuencias. La señal a ruido resultante fue de  $SNR = 124.75$ . La línea negra representa el modelo teórico, utilizando los parámetros presentados en (Bakos et al., 2012). El coeficiente de correlación entre la curva y el modelo es  $C = 0.982$ . La región sombreada representa el egreso del planeta del disco solar, la duración de este evento es  $t_c$ .

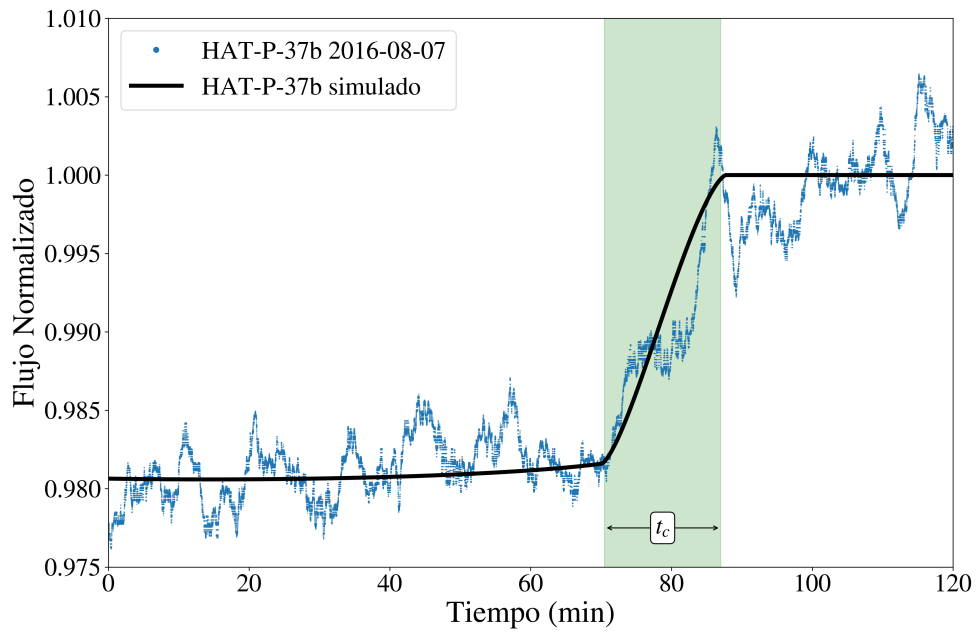


Figura 4.9: Azul: curva de luz de HAT-P-37b después del promedio móvil. La señal a ruido resultante fue de  $SNR = 116.64$ . La línea negra representa el modelo teórico, utilizando los parámetros presentados en (Bakos et al., 2012). El coeficiente de correlación entre la curva y el modelo es  $C = 0.962$ . La región sombreada representa el egreso del planeta del disco solar, la duración de este evento es  $t_c$ .

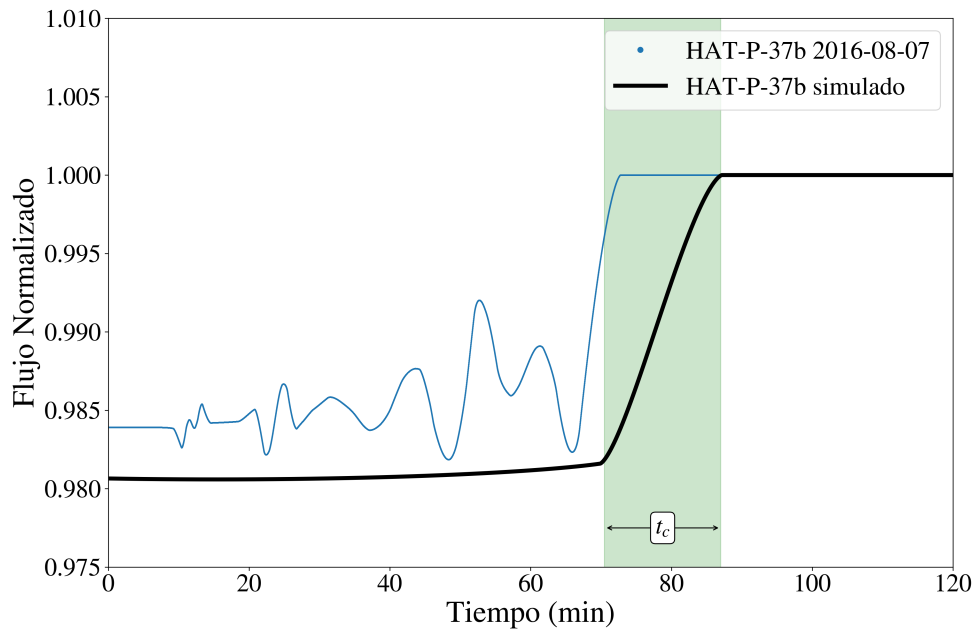


Figura 4.10: Azul: curva de luz de HAT-P-37b después del filtrado con PCA. La señal a ruido resultante fue de  $SNR = 135.51$ . La línea negra representa el modelo teórico, utilizando los parámetros presentados en (Bakos et al., 2012). El coeficiente de correlación entre la curva y el modelo es  $C = 0.889$ . La región sombreada representa el egreso del planeta del disco solar, la duración de este evento es  $t_c$ .

cipales similares a las del modelo, se producirá una curva similar a la que se observa en la figura 4.10. Si la curva es solo ruido, el resultado será una curva errática sin ningún parecido con el modelo como se muestra en la figura 4.11. Por lo tanto, este filtrado nos ayuda a determinar de manera eficiente cuales curvas de luz pueden contener candidatos a tránsitos de exoplanetas.

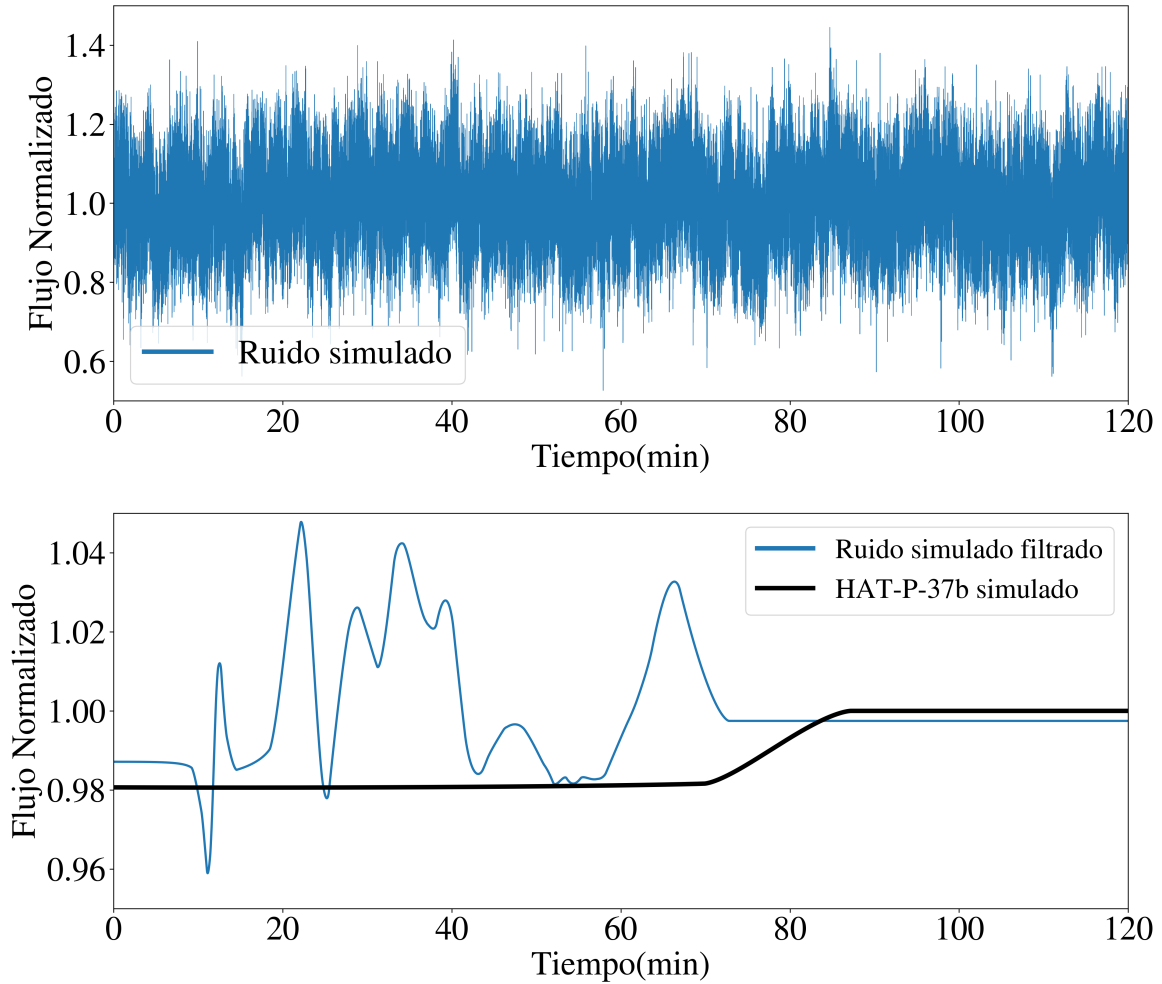


Figura 4.11: Arriba: Ruido simulado con componentes de baja frecuencia en su espectro. Abajo: en azul se muestra el resultado de PCA a una curva sin tránsito inmerso. La línea negra representa el modelo teórico de HAT-P-37b, utilizando los parámetros presentados en (Bakos et al., 2012).

## IV.2.1.4 Resumen de resultados con datos observacionales

En la tabla 4.1 se presentan todos los resultados de la mejora de la SNR utilizando las metodologías descritas anteriormente, para todo el conjunto de curvas de luz reales (véase III.1)

Fecha	Nombre	SNR	SNR <sub>Fourier</sub>	SNR <sub>mov</sub>	SNR <sub>PCA</sub>
06/Ago/2016	WASP-74b <sup>1</sup>	47.03	226.50	212.93	279.51
06/Ago/2016	WASP-74b <sup>2</sup>	39.82	114.47	114.73	156.30
06/Ago/2016	HAT-P-32b	14.55	29.68	29.43	31.85
07/Ago/2016	HAT-P-37 <sup>1</sup>	11.66	176.44	166.00	180.11
07/Ago/2016	HAT-P-37 <sup>2</sup>	11.73	124.75	116.64	135.51
28/May/2017	HAT-P-14b	61.40	154.80	158.96	-
29/May/2017	WASP-74b	9.46	93.33	82.86	93.86
31/May/2017	WASP-48b	7.74	61.25	48.16	63.29
31/May/2017	WASP-74b	14.70	41.36	42.60	47.97
02/Jun/2017	WASP-48b <sup>1</sup>	11.07	28.10	22.91	32.71
02/Jun/2017	WASP-48b <sup>2</sup>	7.67	20.37	14.78	20.32
05/Jun/2017	HAT-P-31b	6.17	68.05	43.67	80.09
08/Jun/2018	Kepler-17b	146.52	523.31	512.25	602.00

Tabla 4.1: Resumen de los resultados para mejorar la SNR en las curvas de luz de tránsitos de exoplanetas obtenidas en múltiples temporadas en el OAN-SPM. SNR indica el valor de ruido después de la fotometría. SNR<sub>Fourier</sub>, SNR<sub>mov</sub> y SNR<sub>PCA</sub> indican el valor de ruido resultado del filtrado de Fourier, promedio móvil y PCA respectivamente. Los superíndices indican que la misma curva se dividió en 2 segmentos de 2 horas cada uno, los cuales se analizaron de manera independiente.

Todas las curvas de luz de la tabla 4.1 tienen una duración de 2 horas y fueron obtenidas al momento del tránsito de exoplanetario. En la mayoría de las ocasiones, solo se observó la entrada o salida del exoplaneta al disco solar, las curvas con el superíndices (<sup>1,2</sup>) indican que se observó el tránsito completo y se separó en segmentos de 2 horas para su análisis, donde (<sup>1</sup> y <sup>2</sup>) indican el comienzo y el final del tránsito respectivamente.

## IV.2.2 Determinación de candidatos a tránsitos

Cabe destacar que las técnicas de filtrado utilizadas para la mejora de la SNR, deforman la curva de luz, por lo que no es posible recuperar la curva exacta del tránsito.

Fecha	Nombre	$\Delta F$ (%)	$\Delta F_{Fourier}$ (%)	$\Delta F_{mov}$ (%)	$\Delta F_{PCA}$ (%)
06/Ago/2016	WASP-74b <sup>1</sup>	1.04	-	-	-
06/Ago/2016	WASP-74b <sup>2</sup>	1.04	-	-	-
06/Ago/2016	HAT-P-32b <sup>1</sup>	2.44	-	-	-
07/Ago/2016	HAT-P-37 <sup>1</sup>	2.04	1.63	1.67	1.75
07/Ago/2016	HAT-P-37 <sup>2</sup>	2.04	1.74	1.84	1.91
28/May/2017	HAT-P-14b <sup>1</sup>	0.5	-	-	-
29/May/2017	WASP-74b <sup>2</sup>	1.04	2.42	2.59	1.62
31/May/2017	WASP-48b <sup>2</sup>	1.0	-	-	-
31/May/2017	WASP-74b <sup>1</sup>	1.04	6.47	6.28	6.59
02/Jun/2017	WASP-48b <sup>1</sup>	1.0	6.63	5.33	7.5
02/Jun/2017	WASP-48b <sup>2</sup>	1.0	-	-	-
05/Jun/2017	HAT-P-31b <sup>1</sup>	1.23	-	-	-
08/Jun/2018	Kepler-17b <sup>1</sup>	2.13	-	-	-

Tabla 4.2: Resultados de  $\Delta F$  obtenidos ajustando el modelo del trapezoide a las curvas de luz de tránsitos.  $\Delta F_{Fourier}$ ,  $\Delta F_{mov}$  y  $\Delta F_{PCA}$  representan el valor de  $\Delta F$  obtenido después del filtrado de Fourier, promedio móvil y PCA respectivamente. El superíndice <sup>1</sup> indica que la curva de luz fue obtenida mediante el inicio del tránsito y <sup>2</sup> indica que la curva contenía el final del tránsito.

Sin embargo, el objetivo es encontrar variaciones en el flujo, para determinar si la estrella observada puede ser candidata a tener un exoplaneta. Por lo que las deformaciones debido al mismo filtrado no afectan el proceso de selección de candidatos.

Se aplicó un ajuste del modelo trapezoidal (véase III.6.1) a las curvas filtradas. El desempeño de los distintos métodos de filtrado, se evaluó utilizando el valor de  $\Delta F$  resultado del ajuste a los datos filtrados y comparándolo con el valor reportado por la comunidad en bases de datos como *Exoplanet Transit Database*. Estos resultados se aprecian en la tabla 4.2.

El ajuste trapezoidal, nos regresa como resultado dos parámetros,  $\Delta F$  y  $t_c$  (véase II.1.1). En estos resultados evaluamos solamente el valor de  $\Delta F$ , esto nos brinda información sobre la magnitud del cambio en el flujo de la estrella observada, el cual es uno de los principales criterios para la selección de un posible candidato a tránsito.

Los signos (-) en la tabla 4.2, indican que el método de ajuste falló. Esto ocurre cuando el ajuste del modelo nos da como resultado un valor de  $\Delta F$  fuera de los límites permitidos. Esto nos indica que la detectabilidad en los datos observacionales fue baja (5

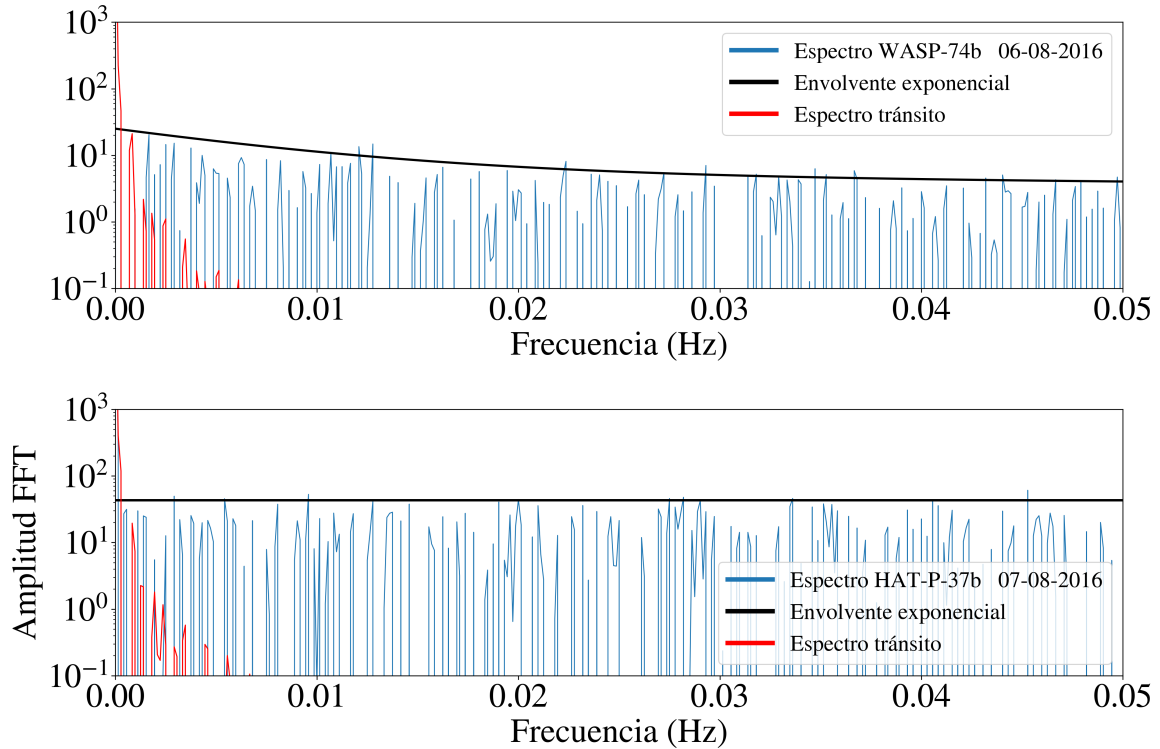


Figura 4.12: Arriba: Espectro de la curva de luz de WASP-74b del 06 de agosto. Abajo: Espectro de la curva de luz de HAT-P-37b del 06 de agosto. La línea roja es el espectro del tránsito simulado utilizando los parámetros presentados en (Bakos et al., 2012). La línea negra representa la envolvente exponencial del espectro.

de 13 eventos), en la mayoría de los casos negativos, las curvas de luz tenían tendencias de baja frecuencia después del proceso de fotometría (véase Anexo 1). En la figura 6.1 comparamos 2 espectros, el primero pertenece a la curva de luz de WASP-74b del 06 de agosto y el de HAT-P-37b del 07 de agosto del 2016.

Comparamos estos espectros, para ilustrar la diferencia entre las curvas en donde el método funciona correctamente (HAT-P-37<sup>2</sup>) y en las que no (WASP-74b<sup>1</sup>) véase tabla 4.2. Podemos apreciar 2 notables diferencias: el nivel de ruido y la envolvente del espectro. El nivel de ruido en la curva de luz HAT-P-37b<sup>2</sup> es mayor que en la de WASP-74b<sup>1</sup>, esto puede observarse en la tabla 4.1. Por lo tanto, la SNR en la curva de luz no juega un papel crucial para que el método funcione. Por otra parte, vemos que la envolvente exponencial del



---

espectro de WASP-74b<sup>1</sup> posee significativas contribuciones de baja frecuencia, mientras que el espectro de HAT-P-37b<sup>2</sup> es completamente plano. Esto también puede apreciarse en la tabla 3.3, existe una correlación entre la magnitud de los parámetros ( $a_1, a_2$ ) y los fallos en la identificación de los tránsitos. Cuando la región de bajas frecuencias (línea roja en 3.9) se encuentra contaminada, es muy difícil recuperar la señal del tránsito inmersa en el ruido. No se encontró un filtrado para remover estas contribuciones de baja frecuencia, por lo que para mejorar los resultados, debemos mejorar la curva de luz desde el proceso de fotometría.

En algunos casos como WASP-48b<sup>1</sup> y HAT-P-31b<sup>1</sup>, las curvas de luz poseen variaciones sistemáticas posiblemente causadas por el movimiento del telescopio. Creemos que en los otros casos fallidos las contribuciones de baja frecuencia fueron causadas al cambio en la masa de aire o al *seeing* de la noche de observación, para mayor detalle véase (Guerrero et al., 2020), en donde se estudia a detalle este fenómeno y utilizan algunas de las imágenes que utilizamos en este trabajo.

## CAPÍTULO V: CONCLUSIONES

En este trabajo se implementaron herramientas de computo para el análisis eficiente de curvas de luz de alta cadencia en la búsqueda de señales débiles y de baja frecuencia, como los tránsitos de exoplanetas. Para la detección de posibles candidatos a tránsitos, se desarrolló un algoritmo que ejecuta dos tareas en curvas de luz de alta cadencia: la mejora de la señal a ruido (SNR) y el criterio para la validación de los posibles candidatos.

Para la simulación de las curvas de luz se llevó a cabo un análisis del ruido en curvas de luz reales que se obtuvieron del telescopio de 0.84 m del OAN-SPM con cadencia de 20 Hz y escala de placa similar a la de TAOS II. Este análisis permitió desarrollar un simulador parametrizado de ruido real, lo cual es la base de las curvas de luz utilizadas en este trabajo.

La simulación de tránsitos se llevó a cabo utilizando las librerías libres de BATMAN que utiliza los modelos de (Mandel and Agol, 2002).

En la literatura, existen diversas técnicas para el suavizado de series de tiempo. Las metodologías presentadas en este trabajo no son las únicas que pueden aplicarse, sin embargo, como vemos en la sección IV, estas tienen un buen desempeño. En promedio, se logró mejorar la SNR en un 500% en curvas de luz reales. En algunos de esos casos ( $\sim 30\%$ ), esto fue suficiente para observar a simple vista la firma del tránsito. En el futuro, se pretende implementar este código como un *pipeline* para generar alarmas de posibles firmas con tránsitos en los datos de TAOS-II, que tiene la ventaja de poder reducir los falsos positivos debido a la observación simultánea y sincronizada de estrellas en los tres telescopios.

Para la validación de posibles candidatos a tránsitos, se utilizó el modelo del trapecoide como primera aproximación a una señal observada debido a un tránsito de exoplaneta. Simplificar esta parte del modelo, nos permite aumentar la eficiencia significativamente sin descuidar la confiabilidad del método. Es importante recalcar que con la metodología

planteada en este trabajo, no se pretende calcular magnitudes físicas precisas. El propósito es analizar una gran cantidad de datos en poco tiempo, y etiquetar posibles estrellas de interés, para en un futuro realizar más observaciones y análisis solo a estos objetivos preseleccionados. Esto es importante en proyectos con grandes bases de datos como *Kepler*, *TESS* y TAOS II donde la cantidad de objetivos es tan grande, que la implementación de pipelines de procesos de búsqueda rápida es indispensable.

Como se mencionó anteriormente, un factor importante que se mantuvo a consideración fue la eficiencia: esta metodología está pensada para implementarse con la base de datos de TAOS II, la cual observará alrededor de  $\sim 50,000$  estrellas por noche, con 3 diferentes telescopios. Esto significa que es necesario tener una metodología que sea capaz de analizar cerca de 150,000 curvas de luz de alta cadencia en menos de 24 horas, esto si queremos evitar el cuello de botella con los datos.

Los campos de observación de TAOS II no tienen traslape con otros proyectos de búsqueda de tránsitos de exoplanetas (a excepción de *TESS* el cual cubre el 90% de la bóveda celeste). Esto es muy importante, debido a que, cualquier detección que logremos en TAOS II, tiene el potencial de llegar a ser un nuevo descubrimiento. Otra ventaja es que TAOS II, cuenta con 3 telescopios observando el mismo campo de observación, esto nos ayuda a tener una mayor confiabilidad en los resultados, porque podremos confirmar el candidato con tres observaciones simultáneas, esto reducirá significativamente la cantidad de falsos positivos. TAOS II realizará observaciones a  $\pm 1$  horas del cenit, lo cual reducirá las tendencias en las curvas de luz, al disminuir el cambio en la masa de aire. También se tendrán telescopios más grandes y mejores detectores que con los que evaluamos la metodología presentada en este trabajo, por lo que esperamos tener mejores resultados con los datos de TAOS II.

Además de la implementación del algoritmo como *pipeline* de TAOS II, un trabajo que ha quedado pendiente para el futuro, es el desarrollo de una metodología para predecir la posible ocurrencia del tránsito una vez que se detecten algunos ingresos o egresos en las curvas de luz. Esto permitiría solicitar tiempo de observación en otros telescopios para dar seguimiento y lograr confirmar la existencia del tránsito.

## ANEXOS

### 6.1. Datos observacionales

En esta sección se presentan todas las curvas de luz, que se obtuvieron durante las distintas temporadas de observación, las cuales se encuentran en la tabla 3.2.

### 6.2. Paquetería HCLcat

HCLcat (del inglés *High Cadence Light curves analytic tools*) es una colección de módulos en Python los cuales se utilizaron para generar los resultados obtenidos en este trabajo. Esta paquetería está en GitHub y se puede instalar utilizando el comando:

```
pip install git+https://github.com/ferwazz/HCLcat.git
```

A continuación se describirán algunos de los módulos disponibles:

- `doble_gaussian(x,a,mu,sigma,a2,mu2,sigma2)`

Modelo de dos gaussianas como (3.2).

- `snr(x)`:

Calcula la SNR de una curva de luz.

- `set_noise_level(curve,noise)`:

Regresa la curva de luz normalizada con SNR igual a `noise`.

- `genruido(noise,lenght=144000)`:

Genera una curva de ruido, con distribución doble gaussiana de longitud `lenght`.

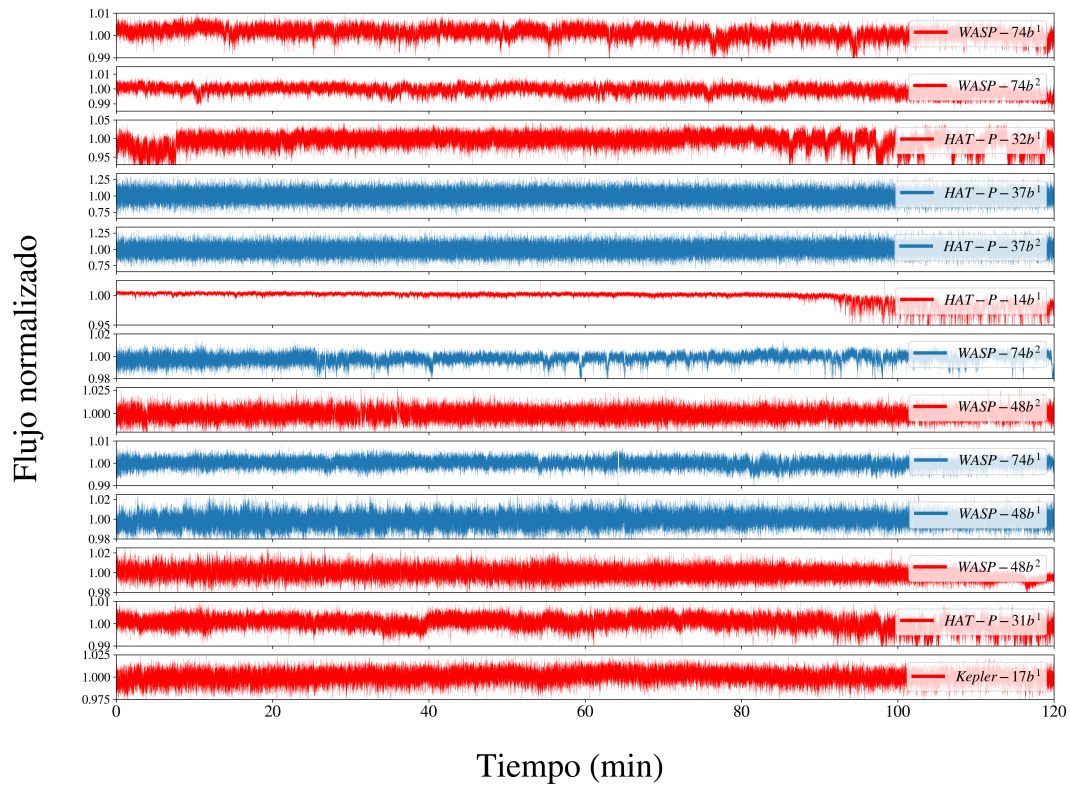


Figura 6.1: Se muestran todas las 13 curvas de luz, a partir de las cuales se encontraron los resultados presentados en la sección IV.2. Las curvas en rojo, son las curvas de luz para las cuales no pudimos detectar el tránsito inmerso en el ruido. Las curvas azules son las observaciones donde si se detectó un tránsito, esto de acuerdo a la tabla 4.1.

- `remove_outliers(x, std_limit=5)`:  
Remueve outliers de una curva `x`, basado en su desviación estándar `std_limit`.
- `prom_mov(x, mov, method=prom)`:  
Aplica un promedio móvil a una curva. `mov` define el tamaño de la ventana de promediado. Si `method=prom` calcula una media de los valores en la ventana, si `method='med'` calcula la mediana.
- `dexp(x, base, a, b, a2, b2)`:  
Modelo de envolvente exponencial para el espectro (3.3).
- `ruido_real(noise, base, a, b, a2, b2, lenght=144000)`:  
Genera una curva de ruido simulado, con componentes de baja frecuencia en su espectro.
- `fourierfilter(x, cutfreq)`:  
Aplica un filtrado pasa bajas de Fourier, dada la frecuencia de corte `cutfreq`.
- `pca_train(data_train)`:  
Calcula las componentes principales de la matriz de datos `data_train`.
- `pca_filter(data_noised, comp)`:  
Dadas las componentes principales `comp`, se reconstruye la curva `data_noised`.
- `medio_trapezoid(x, depth, down, tc)`:  
Modelo del trapecoide (mitad) para ajustar a los posibles candidatos a tránsitos.
- `ruido_real_random(cotas, lenght=144000, snr=0)`:  
Genera una curva de ruido simulado, con parámetros aleatorios dentro de las limites definidos en `cotas`.

Se planea mantener el repositorio y agregar nuevos módulos para el procesamiento de datos de alta cadencia y la búsqueda de características como posibles tránsitos de exoplanetas, variabilidad, etc. A continuación se presentan algunos ejemplos de como

utilizar estos módulos para la implementación de algunas herramientas que utilizamos en este trabajo.

Este código se puede utilizar para generar una base de datos de ruidos simulados. Estos ruidos simulados tienen una envolvente exponencial en su espectro. Los parámetros se eligen de manera aleatoria dentro de las cotas definidas en `bounds`. Al final guarda la base de datos en un archivo con formato HDF5.

```
1 from HCLcat import *
2 import h5py
3
4 signal_noise = [10,20,50,80,100] # SNR deseadas en nuestra base de datos
5 n_curves = 100 # Numero de curvas con la misma SNR pero diferentes
   parametros
6
7 bounds=np.zeros([6,2])
8 bounds[0,0]=10      #SNR MIN
9 bounds[0,1]=100     #SNR MAX
10 bounds[1,0]=0.85   #BASE MIN
11 bounds[1,1]=10     #BASE MAX
12 bounds[2,0]=0      #a MIN
13 bounds[2,1]=30     #a MAX
14 bounds[3,0]=2      #b MIN e-2
15 bounds[3,1]=250    #b MAX e-2
16 bounds[4,0]=0      #a2 MIN
17 bounds[4,1]=150    #a2 MAX
18 bounds[5,0]=2.9    #b2 MIN e-3
19 bounds[5,1]=60     #b2 MAX e-3
20
21 for j in signal_noise:
22     f = h5py.File('database_ruido_SNR_{}.hdf5'.format(j), 'w')
23     for i in range(n_curves):
24         ruido_random,base,a,b,a2,b2 = ruido_real_random(bounds,lenght
=144000,snr=j)
25         dset2 = f.create_dataset(
26             "ruido_{:.3f}_{:.3f}_{:.3f}_{:.3f}_{:.3f}".format(base,a,b,a2,b2),
27             data=ruido_random)
28     f.close()
```



Este código se puede utilizar para generar una base de datos de medios trapecios. Los parámetros se eligen por el usuario. Al final guarda la base de datos en un archivo con formato HDF5.

```
1 from HCLcat import *
2
3 f = h5py.File('database_half_trapezoids.hdf5', 'w')
4
5 delta_f_min = 0.998
6 delta_f_max = 0.94
7 tc_min = 18000
8 tc_max = 72000
9
10 n_radius=25
11 n_tc=25
12
13 depth=np.linspace(delta_f_min, delta_f_max, n_radius)
14 t_caida=np.linspace(tc_min, tc_max, n_tc)
15 xdata=np.linspace(1,144000,144000)
16 down=32000
17
18 for j in np.arange(n_tc):
19     for i in range(n_radius):
20         flux=medio_trapezoid(xdata,depth[i],down,t_caida[j])
21         dset2 = f.create_dataset(
22             "transito_delta_{:.3f}_Tc_{:.3f}".format(depth[i],(t_caida[j]
23             ]/(60*20))),
24             data=flux)
25 f.close()
```

Este código se utilizó para obtener los resultados presentados en este trabajo. Utiliza la base de datos de ruidos simulados, y trapecios previamente generados. La parte de `FILTER_METHOD` puede sustituirse por cualquier método de filtrado que el usuario desee como `prom_mov`, `fourier_filter` ó `pca_filter`. Este código es paralelizable en el sentido de que puede ejecutarse varias veces de manera simultánea y analizar secciones diferentes de las bases de datos.

```
1 from HCLcat import *
2
3 import numpy as np
4 import pickle
5 import h5py
6
7 f=h5py.File('ideal_case_traps.hdf5', 'r')
8
9 size = 625
10 N = 100
11 long = 144000
12
13 depth=np.zeros((size,N))
14 down=np.zeros((size,N))
15 tc=np.zeros((size,N))
16 ruido_num=np.zeros(N)
17 xdata=np.linspace(1,long,long)
18 snr=np.array([10,20,50,80,100])
19
20 tmod = Model(medio_trapezoid)
21 tmod.set_param_hint('tc', value=18000, min=15000, max=73000)
22 tmod.set_param_hint('down', value=np.round(0.008*long))
23 tmod.set_param_hint('depth', value=0.95)
24 params = tmod.make_params()
25
26 for j in snr:
27     f2=h5py.File('database_ruido_SNR_{}.hdf5'.format(j), 'r')
28     keys_ruido = list(f2.keys())
29     for i in range(N):
30         cont=0
31         ruido_num[i] = random.randint(0,99)
```

```
32     ruido = np.array(f2[keys_ruido[int(ruido_num[i])]])
33     for key in f.keys():
34         data = np.array(f[key])
35         data_ruido = data*ruido
36         data_fil = FILTER_METHOD(params)
37         result = tmod.fit(data_fil, params, x=xdata)
38         depth[cont,i] = result.params['depth'].value
39         tc[cont,i] = result.params['tc'].value
40         down[cont,i] = result.params['down'].value
41         cont = cont+1
42     f2.close()
43     pickle.dump(
44     (depth,down,tc),
45     open('param_trap_Fourier_SNR_{}.pickle'.format(j), 'wb'))
46     pickle.dump(
47     (ruido_num),
48     open('ruido_Fourier_SNR_{}.pickle'.format(j), 'wb'))
```

## REFERENCIAS

- Aigrain, S., Barge, P., Deleuil, M., Fressin, F., Moutou, C., Queloz, D., Auvergne, M., Baglin, A., et al. (2007). The corot exoplanet programme: exploring the gas-giant/terrestrial planet transition. *arXiv preprint astro-ph/0702062*.
- Alapini Odunlade, A. E. P. (2010). Transiting exoplanets: characterisation in the presence of stellar activity.
- Allen, C. W. and Cox, A. N. (2001). *Allen's astrophysical quantities*. Springer Science & Business Media.
- Bailey, S. (2012). Principal component analysis with noisy and/or missing data. *Publications of the Astronomical Society of the Pacific*, 124(919):1015.
- Bakos, G., Bayliss, D., Bento, J., Bhatti, W., Brahm, R., Csubry, Z., Espinoza, N., Hartman, J., Henning, T., Jordán, A., et al. (2018). Hats-71b: A giant planet transiting an m3 dwarf star in tess sector 1. *arXiv preprint arXiv:1812.09406*.
- Bakos, G., Csubry, Z., Penev, K., Bayliss, D., Jordán, A., Afonso, C., Hartman, J. D., Henning, T., Kovács, G., Noyes, R. W., et al. (2013). Hatsouth: a global network of fully automated identical wide-field telescopes. *Publications of the Astronomical Society of the Pacific*, 125(924):154.
- Bakos, G., Hartman, J., Torres, G., Béky, B., Latham, D., Buchhave, L. A., Csubry, Z., Kovács, G., Bieryla, A., Quinn, S., et al. (2012). Hat-p-34b-hat-p-37b: Four transiting planets more massive than jupiter orbiting moderately bright stars. *The Astronomical Journal*, 144(1):19.
- Bakos, G., Lazar, J., Papp, I., Sari, P., and Green, E. M. (2002). System description and

- first light curves of the hungarian automated telescope, an autonomous observatory for variability search. *Publications of the Astronomical Society of the Pacific*, 114(799):974.
- Bakos, G., Noyes, R., Kovács, G., Stanek, K., Sasselov, D., and Domsa, I. (2004). Wide-field millimagnitude photometry with the hat: A tool for extrasolar planet detection. *Publications of the Astronomical Society of the Pacific*, 116(817):266.
- Barge, P., Baglin, A., Auvergne, M., et al. (2007). Corot: pioneer space mission for exoplanet transit search. *Proceedings of the International Astronomical Union*, 3(S249):3–16.
- Barge, P., Baglin, A., Auvergne, M., Rauer, H., Léger, A., Schneider, J., Pont, F., Aigrain, S., Almenara, J.-M., Alonso, R., et al. (2008). Transiting exoplanets from the corot space mission-i. corot-exo-1b: a low-density short-period planet around a g0v star. *Astronomy & Astrophysics*, 482(3):L17–L20.
- Borucki, W., Koch, D., Jenkins, J., Sasselov, D., Gilliland, R., Batalha, N., Latham, D., Caldwell, D., Basri, G., Brown, T., et al. (2009). Kepler’s optical phase curve of the exoplanet hat-p-7b. *Science*, 325(5941):709–709.
- Borucki, W. J., Koch, D., Basri, G., Batalha, N., Brown, T., Caldwell, D., Caldwell, J., Christensen-Dalsgaard, J., Cochran, W. D., DeVore, E., et al. (2010). Kepler planet-detection mission: introduction and first results. *Science*, 327(5968):977–980.
- Borucki, W. J., Koch, D. G., Batalha, N., Bryson, S. T., Rowe, J., Fressin, F., Torres, G., Caldwell, D. A., Christensen-Dalsgaard, J., Cochran, W. D., et al. (2012). Kepler-22b: A 2.4 earth-radius planet in the habitable zone of a sun-like star. *The Astrophysical Journal*, 745(2):120.
- Brault, J. and White, O. (1971). The analysis and restoration of astronomical data via the fast fourier transform. *Astronomy and Astrophysics*, 13:169.
- Bryson, S. T., Batalha, N. M., Colon, K. D., Coughlin, J. L., Haas, M. R., Henze, C., Huber, D., Morton, T., Rowe, J. F., Mullally, S. E., et al. (2017). Kepler certified false positive table.
- Butler, R. P. and Marcy, G. W. (1996). A planet orbiting 47 ursae majoris. *The Astrophysical Journal Letters*, 464(2):L153.

- Cameron, A. C., Bouchy, F., Hébrard, G., Maxted, P., Pollacco, D., Pont, F., Skillen, I., Smalley, B., Street, R., West, R., et al. (2007). Wasp-1b and wasp-2b: two new transiting exoplanets detected with superwasp and sophie. *Monthly Notices of the Royal Astronomical Society*, 375(3):951–957.
- Charbonneau, D., Brown, T. M., Latham, D. W., and Mayor, M. (1999). Detection of planetary transits across a sun-like star. *The Astrophysical Journal Letters*, 529(1):L45.
- Charpinet, S., Fontaine, G., Brassard, P., Green, E. M., Van Grootel, V., Randall, S. K., Silvotti, R., Baran, A., Østensen, R., Kawaler, S. D., et al. (2011). A compact system of small planets around a former red-giant star. *Nature*, 480(7378):496.
- Chong, E. K. and Zak, S. H. (2004). *An introduction to optimization*. John Wiley & Sons.
- Chromey, F. (1996). Special considerations for flat fielding. *CCD Astronomy*, pages 18–23.
- Collier Cameron, A., Pollacco, D., Street, R., Lister, T., West, R., Wilson, D., Pont, F., Christian, D., Clarkson, W., Enoch, B., et al. (2006). A fast hybrid algorithm for exoplanetary transit searches. *Monthly Notices of the Royal Astronomical Society*, 373(2):799–810.
- Cooley, J. W. and Tukey, J. W. (1965). An algorithm for the machine calculation of complex fourier series. *Mathematics of computation*, 19(90):297–301.
- Désert, J.-M., Charbonneau, D., Demory, B.-O., Ballard, S., Carter, J. A., Fortney, J. J., Cochran, W. D., Endl, M., Quinn, S. N., Isaacson, H. T., et al. (2011). The hot-jupiter kepler-17b: Discovery, obliquity from stroboscopic starspots, and atmospheric characterization. *The Astrophysical Journal Supplement Series*, 197(1):14.
- Enoch, B., Anderson, D., Barros, S., Brown, D., Cameron, A. C., Faedi, F., Gillon, M., Hébrard, G., Lister, T., Queloz, D., et al. (2011). Wasp-35b, wasp-48b, and hat-p-30b/wasp-51b: two new planets and an independent discovery of a hat planet. *The Astronomical Journal*, 142(3):86.
- Fressin, F., Torres, G., Charbonneau, D., Bryson, S. T., Christiansen, J., Dressing, C. D., Jenkins, J. M., Walkowicz, L. M., and Batalha, N. M. (2013). The false positive rate of kepler and the occurrence of planets. *The Astrophysical Journal*, 766(2):81.

- Gilbert, E. A., Barclay, T., Schlieder, J. E., Quintana, E. V., Hord, B. J., Kostov, V. B., Lopez, E. D., Rowe, J. F., Hoffman, K., Walkowicz, L. M., et al. (2020). The first habitable zone earth-sized planet from tess. i: Validation of the toi-700 system. *arXiv preprint arXiv:2001.00952*.
- Golub, G. H. and Reinsch, C. (1971). Singular value decomposition and least squares solutions. In *Linear Algebra*, pages 134–151. Springer.
- Guerrero, C., Hernández-Águila, J., Castro-Chacón, J., Hernández-Valencia, B., Sánchez, E., Reyes-Ruiz, M., Silva, J., and Álvarez-Santana, F. (2020). Effects of atmospheric turbulence on high cadence observations of stellar occultations. *Publications of the Astronomical Society of the Pacific*, 132(1012):064501.
- Hartman, J. D., Bakos, G., Torres, G., Latham, D. W., Kovács, G., Béky, B., Quinn, S., Mazeh, T., Shporer, A., Marcy, G., et al. (2011). Hat-p-32b and hat-p-33b: Two highly inflated hot jupiters transiting high-jitter stars. *The Astrophysical Journal*, 742(1):59.
- Hatzes, A. P., Cochran, W. D., Endl, M., McArthur, B., Paulson, D. B., Walker, G. A., Campbell, B., and Yang, S. (2003). A planetary companion to  $\gamma$  cephei a. *The Astrophysical Journal*, 599(2):1383.
- Hellier, C., Anderson, D., Cameron, A. C., Delrez, L., Gillon, M., Jehin, E., Lendl, M., Maxted, P., Pepe, F., Pollacco, D., et al. (2015). Three wasp-south transiting exoplanets: Wasp-74b, wasp-83b, and wasp-89b. *The Astronomical Journal*, 150(1):18.
- Hippke, M. and Heller, R. (2019). Optimized transit detection algorithm to search for periodic transits of small planets. *Astronomy & Astrophysics*, 623:A39.
- Jenkins, J. M., Caldwell, D. A., Chandrasekaran, H., Twicken, J. D., Bryson, S. T., Quintana, E. V., Clarke, B. D., Li, J., Allen, C., Tenenbaum, P., et al. (2010a). Overview of the kepler science processing pipeline. *The Astrophysical Journal Letters*, 713(2):L87.
- Jenkins, J. M., Chandrasekaran, H., McCauliff, S. D., Caldwell, D. A., Tenenbaum, P., Li, J., Klaus, T. C., Cote, M. T., and Middour, C. (2010b). Transiting planet search in the kepler pipeline. In *Software and Cyberinfrastructure for Astronomy*, volume 7740, page 77400D. International Society for Optics and Photonics.

- Johns-Krull, C. M., McCullough, P. R., Burke, C. J., Valenti, J. A., Janes, K., Heasley, J., Prato, L., Bissinger, R., Fleenor, M., Foote, C., et al. (2008). Xo-3b: a massive planet in an eccentric orbit transiting an f5 v star. *The Astrophysical Journal*, 677(1):657.
- Kipping, D. M., Hartman, J., Bakos, G. Á., Torres, G., Latham, D. W., Bayliss, D., Kiss, L. L., Sato, B., Béky, B., Kovács, G., et al. (2011). Hat-p-31b, c: A transiting, eccentric, hot jupiter and a long-period, massive third body. *The Astronomical Journal*, 142(3):95.
- Kipping, D. M. and Sandford, E. (2016). Observational biases for transiting planets. *Monthly Notices of the Royal Astronomical Society*, 463(2):1323–1331.
- Klückers, V., Woeder, N., Dainty, J., and Longmore, A. (1996). Comparison of shift-and-add and bispectrum image reconstruction methods for astronomy in the near infrared. *JOSA A*, 13(7):1577–1583.
- Kosarev, E. and Pantos, E. (1983). Optimal smoothing of 'noisy' data by fast fourier transform. *Journal of Physics E: Scientific Instruments*, 16(6):537.
- Kovács, G., Zucker, S., and Mazeh, T. (2002). A box-fitting algorithm in the search for periodic transits. *Astronomy & Astrophysics*, 391(1):369–377.
- Kreidberg, L. (2015). batman: Basic transit model calculation in python. *Publications of the Astronomical Society of the Pacific*, 127(957):1161.
- Léger, A., Rouan, D., Schneider, J., Barge, P., Fridlund, M., Samuel, B., Ollivier, M., Guenther, E., Deleuil, M., Deeg, H., et al. (2009). Transiting exoplanets from the corot space mission-viii. corot-7b: the first super-earth with measured radius. *Astronomy & Astrophysics*, 506(1):287–302.
- Lehner, M., Coehlo, N., Zhang, Z.-W., Bianco, F., Wang, J.-H., Rice, J., Protopapas, P., Alcock, C., Axelrod, T., Byun, Y.-I., et al. (2010). The taos project: statistical analysis of multi-telescope time series data. *Publications of the Astronomical Society of the Pacific*, 122(894):959.
- Lehner, M. J., Wang, S.-Y., Alcock, C. A., Cook, K. H., Furesz, G., Geary, J. C., Hiriart, D., Ho, P. T., Lee, W., Melsheimer, F., et al. (2012). The transneptunian automated



- occultation survey (taos ii). In *Ground-based and Airborne Telescopes IV*, volume 8444, page 84440D. International Society for Optics and Photonics.
- Lehner, M. J., Wang, S.-Y., Reyes-Ruiz, M., Alcock, C., Castro, J., Chen, W.-P., Chu, Y.-H., Cook, K. H., Figueroa, L., Geary, J. C., et al. (2016). Status of the transneptunian automated occultation survey (taos ii). In *Ground-based and Airborne Telescopes VI*, volume 9906, page 99065M. International Society for Optics and Photonics.
- Luisier, F., Blu, T., and Unser, M. (2010). Image denoising in mixed poisson–gaussian noise. *IEEE Transactions on image processing*, 20(3):696–708.
- Luu, J. X. and Jewitt, D. C. (2002). Kuiper belt objects: Relics from the accretion disk of the sun. *Annual Review of Astronomy and Astrophysics*, 40(1):63–101.
- Mahu, R., Rojo, P., Firoozabadi, A. D., Soto, I., Sedaghati, E., and Yoma, N. B. (2017). Estimation of exoplanetary planet-to-star radius ratio with homomorphic processing. *Astronomy and computing*, 20:160–167.
- Mandel, K. and Agol, E. (2002). Analytic light curves for planetary transit searches. *The Astrophysical Journal Letters*, 580(2):L171.
- Marcy, G., Butler, R., and Fischer, D. (1999). Three jupiter-mass companions orbiting upsilon andromedae. In *Bulletin of the American Astronomical Society*, volume 31, page 847.
- Marcy, G. W. and Butler, R. P. (1996). A planetary companion to 70 virginis. *The Astrophysical Journal*, 464(2):L147–L151.
- Mayor, M. and Queloz, D. (1995). A jupiter-mass companion to a solar-type star. *Nature*, 378(6555):355.
- McGurk, R. C., Kimball, A. E., and Ivezić, Ž. (2010). Principal component analysis of sloan digital sky survey stellar spectra. *The Astronomical Journal*, 139(3):1261.
- Medeiros, L., Lauer, T. R., Psaltis, D., and Özel, F. (2018). Principal component analysis as a tool for characterizing black hole images and variability. *The Astrophysical Journal*, 864(1):7.

- Mortara, L. and Fowler, A. (1981). Evaluations of charge-coupled device (ccd) performance for astronomical use. In *Solid-State Imagers for Astronomy*, volume 290, pages 28–33. International Society for Optics and Photonics.
- Morton, T. D. (2012). An efficient automated validation procedure for exoplanet transit candidates. *The Astrophysical Journal*, 761(1):6.
- Moutou, C., Deleuil, M., Guillot, T., Baglin, A., Bordé, P., Bouchy, F., Cabrera, J., Csizmadia, S., Deeg, H. J., et al. (2013). Corot: Harvest of the exoplanet program. *Icarus*, 226(2):1625–1634.
- Orosz, J. A., Welsh, W. F., Carter, J. A., Fabrycky, D. C., Cochran, W. D., Endl, M., Ford, E. B., Haghighipour, N., MacQueen, P. J., Mazeh, T., et al. (2012). Kepler-47: a transiting circumbinary multiplanet system. *Science*, 337(6101):1511–1514.
- Owen, M. (2007). *Practical signal processing*. Cambridge university press.
- Pollacco, D., Skillen, I., Collier Cameron, A., Loeillet, B., Stempels, H., Bouchy, F., Gibson, N., Hebb, L., Hébrard, G., Joshi, Y., et al. (2008). Wasp-3b: a strongly irradiated transiting gas-giant planet. *Monthly Notices of the Royal Astronomical Society*, 385(3):1576–1584.
- Pollacco, D. L., Skillen, I., Cameron, A. C., Christian, D. J., Hellier, C., Irwin, J., Lister, T., Street, R., West, R. G., Anderson, D., et al. (2006). The wasp project and the super-wasp cameras. *Publications of the Astronomical Society of the Pacific*, 118(848):1407.
- Pont, F., Zucker, S., and Queloz, D. (2006). The effect of red noise on planetary transit detection. *Monthly Notices of the Royal Astronomical Society*, 373(1):231–242.
- Ricker, G. R., Winn, J. N., Vanderspek, R., Latham, D. W., Bakos, G. Á., Bean, J. L., Berta-Thompson, Z. K., Brown, T. M., Buchhave, L., Butler, N. R., et al. (2014). Transiting exoplanet survey satellite. *Journal of Astronomical Telescopes, Instruments, and Systems*, 1(1):014003.
- Romanishin, W. (2006). An introduction to astronomical photometry using ccds. *University of Oklahoma*, 17.

- Sánchez, E., Castro-Chacón, J., Silva, J., Hernández-Águila, J., Reyes-Ruiz, M., Guerrero, C., Hernández-Valencia, B., Alvarez-Santana, F., García-Díaz, M., et al. (2019). Apphi: An automated photometry pipeline for high cadence, large volume data. *Astronomy and computing*, 26:1–13.
- Schliegel, J. (2017). Tess observatory guide, version 1.1. *TESS Guest Investigator Program: TESS Observatory Guide, Version 1.1 (June 30, 2017)*. TESS Science Support Center, NASA Goddard Space Flight Center, Greenbelt, MD.
- Shallue, C. J. and Vanderburg, A. (2018). Identifying exoplanets with deep learning: A five-planet resonant chain around kepler-80 and an eighth planet around kepler-90. *The Astronomical Journal*, 155(2):94.
- Shin, K., Hammond, J., and White, P. (1999). Iterative svd method for noise reduction of low-dimensional chaotic time series. *Mechanical Systems and Signal Processing*, 13(1):115–124.
- Sullivan, P. W., Winn, J. N., Berta-Thompson, Z. K., Charbonneau, D., Deming, D., Dressing, C. D., Latham, D. W., Levine, A. M., McCullough, P. R., Morton, T., et al. (2015). The transiting exoplanet survey satellite: simulations of planet detections and astrophysical false positives. *The Astrophysical Journal*, 809(1):77.
- Thompson, R. O. (1983). Low-pass filters to suppress inertial and tidal frequencies. *Journal of Physical Oceanography*, 13(6):1077–1083.
- Torres, G., Bakos, G., Hartman, J., Kovács, G., Noyes, R., Latham, D., Fischer, D., Johnson, J., Marcy, G., Howard, A., et al. (2010). Hat-p-14b: A 2.2 mj exoplanet transiting a bright f star. *The Astrophysical Journal*, 715(1):458.
- Vogt, S. S., Marcy, G. W., Butler, R. P., and Apps, K. (2000). Six new planets from the keck precision velocity survey. *The Astrophysical Journal*, 536(2):902.
- Walker, G. A. (2012). The first high-precision radial velocity search for extra-solar planets. *New Astronomy Reviews*, 56(1):9–15.
- Winn, J. N. (2010). Transits and occultations. *arXiv preprint arXiv:1001.2010*.