



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE CIENCIAS

ANÁLISIS DE ESTABILIDAD DE PARTÍCULAS
NEUTRAS DETECTADAS POR EL TELESCOPIO DE
NEUTRONES SOLARES EN SIERRA NEGRA,
PUEBLA

T E S I S

QUE PARA OBTENER EL TÍTULO DE:
Matemática

PRESENTA:

Griselda Barón Martínez

TUTOR:

Dr. Luis Xavier González Méndez



Ciudad Universitaria, 2017

Índice general

Resumen	vii
Introducción	ix
1. Los rayos cósmicos	1
1.1. ¿Qué son los rayos cósmicos?	1
2. El TNS y la red mundial de observatorios	7
2.1. Red mundial de TNS	7
2.2. El TNS en Sierra negra	9
2.3. Datos del TNS en Sierra Negra	11
2.4. Estructura de archivos registrados	13
3. R y desarrollo de códigos	17
3.1. Entorno de programación R	18
3.2. Limpieza y ordenamiento de datos	21
3.3. Paquetes usados de R	24
3.3.1. dplyr	25
3.3.2. lubridate	25
3.3.3. Openair	26
4. Análisis de estabilidad	29
4.1. Introducción	29
4.2. Estabilidad de la variación diurna	30
4.3. Señal total del TNS	54
Conclusiones	57
A. Fulguraciones solares	59
Referencias	63

Índice de figuras

1.1. Espectro de energía de rayos cósmicos.	2
1.2. Observaciones mensuales del número manchas solares (Solar Influences Data Analysis Center (SIDC)) desde abril de 1964, y medias mensuales de las cuentas del monitor de neutrones de Oulu (Finlandia), ajustados por presión barométrica y eficiencia. Actualización: 1 de marzo de 2017.	4
1.3. Decrecimiento Forbush detectado por el monitor de neutrones de la Ciudad de México el 22 de junio de 2015.	5
2.1. Localización geográfica de los TNS alrededor del planeta.	8
2.2. Esquema del TNS instalado en la cima del volcán Sierra Negra, Puebla.	10
2.3. Gráfica que muestra el número de días registrados por mes y año.	13
3.1. Entrada de datos en R	19
3.2. Creando listas y dataframes.	20
4.1. Variación diurna registrada por el TNS en Sierra Negra. 3 de febrero de 2014 .	30
4.2. Variación temporal de partículas neutras para el canal <i>S1_A</i> en 2004	31
4.3. Variación de las cuentas en 2004 para los cuatro canales	32
4.4. Variación de las cuentas para <i>S1_A</i> en 2005	33
4.5. Variación de las cuentas en 2005 para los cuatro canales	34
4.6. Variación de las cuentas para <i>S1_A</i> en 2006	34
4.7. Variación diurna en 2006 para los cuatro canales con anti	35
4.8. Variación de las cuentas para <i>S1_A</i> en 2008	36
4.9. Variación diurna en 2008 para los cuatro canales con anti	37
4.10. Variación de las cuentas para <i>S1_A</i> en 2009	38
4.11. Variación diurna en 2009 para los cuatro canales con anti	39
4.12. Variación de las cuentas para <i>S1_A</i> en 2010	40
4.13. Variación diurna en 2010 para los cuatro canales con anti	41
4.14. Variación de las cuentas para <i>S1_A</i> en 2011	42
4.15. Variación de las cuentas para los cuatro canales en 2011	43
4.16. Variación de las cuentas para <i>S1_A</i> en 2012	44
4.17. Variación diurna en 2012 para los cuatro canales con anti	45
4.18. Variación de las cuentas para <i>S1_A</i> en 2013	46
4.19. Variación diurna en 2013 para los cuatro canales con anti	47

4.20. Variación de las cuentas para <i>S1_A</i> en 2014	48
4.21. Variación diurna en 2014 para los cuatro canales con anti	49
4.22. Variación de las cuentas para <i>S1_A</i> en 2015	50
4.23. Variación diurna en 2015 para los cuatro canales con anti	51
4.24. Variación de las cuentas para <i>S1_A</i> en 2016	51
4.25. Comportamiento de los máximos de la variación diurna	53
4.26. Comportamiento por año de los máximos de la variación diurna	53
4.27. Datos para el canal <i>S1_A</i> con razón de conteo 1 min.	55
4.28. Señal completa para <i>S1_A</i> con una razón de conteo de 24 horas.	56
A.1. Fulguración solar de clase M, 12 de enero de 2015	59
A.2. Fulguración solar de clase X, 27 de enero de 2012	60
A.3. El Sol con baja actividad	61

Índice de tablas

2.1. Datos de la Red Mundial de TNS.	9
2.2. Datos del 20 de julio de 2004, razón de conteo de 1 minuto.	14
2.3. Datos del 1 de junio de 2008, razón de conteo 10 segundos.	15
2.4. Datos del 10 de marzo de 2008, razón de conteo 3 segundos.	16
4.1. Parámetros característicos para los datos del año 2004	32
4.2. Parámetros característicos para los datos del año 2005	33
4.3. Parámetros característicos para los datos del año 2006	35
4.4. Parámetros característicos para los datos del año 2008	37
4.5. Parámetros característicos para los datos del año 2009	38
4.6. Parámetros característicos para los datos del año 2010	40
4.7. Parámetros característicos para los datos del año 2011	42
4.8. Parámetros característicos para los datos del año 2012	44
4.9. Parámetros característicos para los datos de 2013	46
4.10. Parámetros característicos para los datos del año 2014	48
4.11. Parámetros característicos para los datos del año 2015	50
4.12. Parámetros característicos para los datos del año 2016	52

Resumen

El Telescopio de Neutrones Solares (TNS) de México se encuentra instalado en la cima del volcán Sierra Negra, Puebla, a 4580 m s.n.m. y está operando desde el año 2004. El TNS cuenta con ocho canales de deposición de energía (E) de partículas incidentes que corresponden a $E \geq 30 \text{ MeV}$, 60 MeV , 90 MeV y 120 MeV , cuatro de éstos canales corresponden a partículas cargadas y los otros cuatro a partículas neutras. Además de medir el fondo de rayos cósmicos galácticos, el TNS tiene la capacidad de detectar la energía y dirección de arribo del flujo de neutrones solares.

La base de datos del TNS cuenta hasta la fecha con 13 años de información sobre la detección de rayos cósmicos. En este trabajo se presenta un análisis detallado de estabilidad estadística para la serie total de datos registrados en los canales de partículas neutras desde el año 2004 hasta el año 2016, con una razón de conteo de 10 segundos. Para este análisis se utilizó el software estadístico R, donde se muestran los principales códigos ejecutados.

Una vez que se tienen los datos limpios y ordenados se procede con el análisis estadístico y gráfico para después pasar a la interpretación del análisis. Así, este trabajo muestra las variaciones que se han presentado en los registros de partículas neutras para conocer la calidad de los datos detectados y las distintas influencias electrónicas, eléctricas y de fenómenos de actividad atmosférica que han generado cambios en la estadística anual del TNS. Se da a conocer la señal total del detector y cómo ha variado a lo largo del tiempo. De esta manera, el análisis nos permite asegurar que se está trabajando con datos confiables para realizar los estudios básicos de física solar y poder conocer y diferenciar las afectaciones de los fenómenos puramente eléctrico-electrónicos y/o de otra índole.

Introducción

La superficie de nuestro planeta es bombardeada continuamente por pequeñas y misteriosas partículas que viajan a través del espacio. Estas partículas son los rayos cósmicos que provienen del Sol y de fuera del Sistema Solar. Desde su descubrimiento en 1912 por el físico austriaco Victor Franz Hess¹ y hasta el descubrimiento del antiproton por un acelerador de partículas en 1955, la radiación cósmica ha sido el instrumento científico más importante para avanzar en el estudio de las propiedades de las partículas subatómicas [2].

Los rayos cósmicos son principalmente núcleos atómicos, despojados de sus electrones por los procesos de aceleración en la interacción desde la fuente hasta la Tierra. Hay tres tipos de rayos cósmicos: los galácticos, anómalos y solares. Los *rayos cósmicos galácticos* son de energía muy alta (hasta 10^{20} eV) y la mayor parte de ésta se genera a partir del nacimiento de las supernovas. Los *rayos cósmicos anómalos* se originan en el medio interplanetario y se denominan así porque su composición es inusual; no sigue las abundancias naturales predichas para los diferentes isótopos² [21]. Hoy en día se sabe que en la atmósfera solar también se producen una gran cantidad de partículas a las que llamamos *rayos cósmicos solares* y cuya energía puede ser del orden de 10^{10} eV^3 [3]. El Sol, además de ser nuestra principal fuente de luz y calor, es el mayor acelerador de partículas del Sistema Solar, el cual es capaz de producir neutrones relativistas, principalmente en fulguraciones clase X (véase el apéndice A).

Una fulguración solar es una explosión en el Sol que ocurre cuando la energía almacenada en campos magnéticos intensos se libera repentinamente. Esta explosión genera muchas veces una gran cantidad de partículas de energía muy alta, la mayor parte son partículas cargadas (electrones, protones y núcleos de elementos más pesados), éstas son desviadas por campos magnéticos en el sitio de aceleración, en el medio interplanetario y en el campo geomagnético. De esta forma, al llegar a la Tierra, la información física relevante se pierde. Las partículas neutras, al no ser afectadas por el campo magnético interplanetario, conservan información relevante para comprender y calcular el espectro de los iones y su tiempo de producción. [3].

Los neutrones solares contienen información importante sobre los mecanismos de aceleración de iones en la atmósfera del Sol [26]. Con base en que el tiempo de vida media de los

¹Galardonado con el Premio Nobel de Física en 1936.

²Núcleos del mismo elemento que tienen diferente número de neutrones.

³Los *eV* (electrón-volts) son unidades que miden la energía cinética adquirida por un electrón cuando es acelerado en un campo eléctrico producido por una diferencia de potencial de un volt. Ver más en [15] y [19].

neutrones libres es de 886 segundos, tienen que ser relativistas¹ para que sean detectados en tierra.

El mecanismo de aceleración de los electrones ha sido estudiado mediante las observaciones con rayos X y rayos gamma (rayos γ), mientras que el de los iones aún no se comprende totalmente [3]. La información de la aceleración de iones se transfiere a los neutrones solares y las líneas de rayos γ que son generados por las reacciones nucleares de las partículas aceleradas con núcleos en la atmósfera solar [8]. De aquí la importancia de detectar neutrones solares.

Es necesario conocer la energía de los neutrones solares para calcular el tiempo de aceleración y saber si la aceleración de partículas es gradual o impulsiva, para ello se debe tener registro simultáneo de neutrones solares en tierra y de las emisiones rayos X y rayos γ en el espacio [3].

En el instituto de Geofísica de la UNAM se detectan rayos cósmicos con energías desde los 8.2 GeV² con un instrumento que se conoce como Monitor de Neutrones³ (MN) [9]. El MN mide el flujo de rayos cósmicos galácticos y las variaciones debidas a las emisiones de la actividad del Sol, además de rayos cósmicos solares que se emiten cuando hay fulguraciones extremas. El MN tiene una alta sensibilidad, pero no puede conocer la dirección de arriba y energía de las partículas incidentes y no discrimina entre protones y neutrones. Por esta razón, es necesario usar un equipo nuevo especializado en la detección de neutrones solares.

El Telescopio de Neutrones Solares (TNS) es capaz de medir el flujo de neutrones solares, la energía primaria y diferenciar entre neutrones y protones. El nombre de telescopio se debe a que además de la capacidad mencionada anteriormente, mide la dirección de arriba de las partículas incidentes. Existe una red mundial de TNS instalados en altas montañas (Armenia, Bolivia, China, Hawái, Japón, Suiza y México). En México se instaló en la cima del volcán Sierra Negra, Puebla, en colaboración con el STELab (Solar Terrestrial Environment Laboratory) de la Universidad de Nagoya, Japón [25]. Ha estado funcionando de manera casi continua desde julio de 2004 y ha registrado dos eventos de neutrones solares producidos en la fulguración del 7 de septiembre de 2005 y el 8 de julio de 2014 [4, 13].

Desde 2004 hasta 2016 se tienen 13 años de registro de la detección del TNS con cuatro canales de deposición de energía. Debido al interés por observar y comprender los fenómenos y mecanismos físicos que produce el Sol, es de gran importancia hacer un análisis estadístico detallado y estudiar la estabilidad de los datos a lo largo del tiempo antes de hacer estudios de física solar. El análisis se realiza para datos de los cuatro canales de energía que registran partículas neutras, con el uso del software estadístico R.

Este trabajo está estructurado en 4 capítulos. En el capítulo 1 se da una breve introducción a los rayos cósmicos, en particular a los rayos cósmicos solares y se explican las principales variaciones observadas.

¹ $E > m_o c^2$ donde m_o es la masa en reposo de la partícula y c es la velocidad de la luz en el vacío.

²Rigidez umbral. Véase el *capítulo 1*

³El MN instalado en C.U. es considerado uno de los detectores de neutrones más estables del mundo [9].

En el capítulo 2 se muestra la red mundial de Telescopios de Neutrones Solares; en particular las características del TNS instalado en Sierra Negra, así como el diseño, canales y capacidad de detección. También se dan las características de los datos registrados por el TNS en Sierra Negra, ya que durante los 13 de años de detección se cambió la programación del registro de datos debido a ajustes y calibración del equipo.

Con el avance de la tecnología, el proceso de adquisición de datos se ha vuelto más eficiente, esto implica gran cantidad de almacenamiento de datos. Para manipular, analizar y extraer información importante de estos datos es necesario hacer uso de algún software para programación estadística. Por esta razón, el capítulo 3 se dedica al lenguaje R, el cual se usó para llevar a cabo el análisis estadístico del presente trabajo. Se muestran los códigos programados; desde la importación y limpieza de los datos del TNS, hasta el análisis gráfico y estadístico.

El capítulo 4 muestra un análisis detallado de las distintas variaciones que ha presentado el flujo de partículas neutras durante 13 años para los cuatro canales de energía con anticoincidencia electrónica.

Finalmente se exponen las conclusiones del análisis estadístico y se muestra la señal total del detector (2004-2016).

Capítulo 1

Los rayos cósmicos

1.1. ¿Qué son los rayos cósmicos?

Los rayos cósmicos son núcleos ionizados mediante procesos de aceleración que viajan a través de la Vía Láctea, incluyendo el sistema solar. Se producen en estrellas, remanentes de supernovas, núcleos activos de galaxias y otros objetos astrofísicos. La composición de los rayos cósmicos es de aprox 93 % de protones, 6.3 % de partículas alfa y 0.7% son núcleos de elementos más pesados. La pregunta fundamental es, ¿dónde se originan? y en particular, ¿como es que son aceleradas a tan altas energías?. La respuesta a la pregunta del origen de los rayos cósmicos todavía no es totalmente conocido. Es claro, sin embargo, que casi todos ellos vienen desde fuera del sistema solar, pero desde dentro de la galaxia [?].

Se denominan rayos cósmicos primarios a las partículas que inciden en la atmósfera terrestre procedentes del espacio, compuestos principalmente de protones y partículas alfa. Al interactuar con un núcleo atmosférico producen una cascada de partículas compuesta por rayos γ , electrones, muones,etc, que son llamados rayos cósmicos secundarios [?].

El *espectro de energía* de la radiación cósmica primaria describe cómo están distribuidas con respecto a la energía. En la Figura 1.1

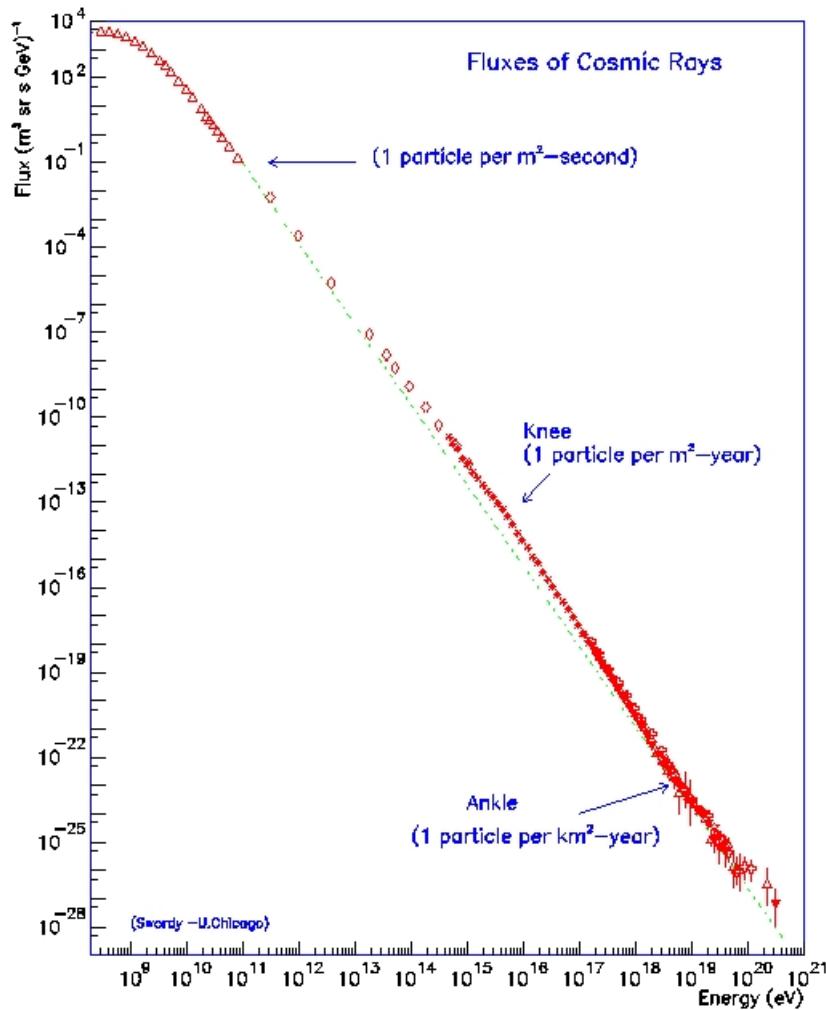


Figura 1.1: Espectro de energía de rayos cósmicos.

Existe un gran número de mediciones y estudios del sector del espectro con energías menores al tobillo; la existencia de la rodilla está aún incompletamente explicada. Para energías mayores al tobillo los rayos cósmicos ya no pueden ser confinados por el campo magnético de la Vía Láctea y provienen de todas las direcciones por igual. Esto indica que estos rayos cósmicos proceden de fuera de la Galaxia. Además, las mediciones del grupo experimental Fly's Eye sugieren que los rayos bajo el tobillo son principalmente núcleos de átomos pesados(mayormente Fe), mientras que los eventos sobre el tobillo son exclusivamente protones y neutrones. Esto es consistente con la hipótesis de que los rayos supra-tobillos son extragalácticos, ya que los núcleos pesados con energías superiores a 4×10^{18} eV no podrían sobrevivir un viaje extragaláctico debido al roce con el Fondo de Radiación Cósmica (FRC), mecanismo que se describirá más adelante.

La Heliosfera es responsable de afectar a las partículas con energías que van desde los 10^9 eV a los 10^{15} eV, en la región del espectro conocida como “rodilla”, por debajo de esta zona,

los RC galácticos son la contribución dominante. Las partículas con energías de $\tilde{1}015\text{-}1017$ eV son probablemente aceleradas por mecanismos de choques rápidos como los remanentes de supernovas. Las partículas con energías superiores son llamadas “de ultra alta energía”, poseen radios de giro comparables con el radio de nuestra galaxia, son muy poco abundantes en comparación a los RC galácticos y su fuente aún no está plenamente confirmada, debido a esto, a menudo se les considera RC extra-galácticos. Alrededor de 1020 eV, el espectro comienza a aplanarse de nuevo, formando la región conocida como “tobillo”.

Durante muchos años se ha observado que la intensidad de los rayos cósmicos medida en la Tierra experimenta variaciones periódicas en función a las emisiones de la actividad del Sol, es decir, estas variaciones son producidas por la *modulación solar* [5]. La actividad solar modula el flujo de rayos cósmicos detectados en tierra. Entre las variaciones observadas en la radiación cósmica y que son influenciadas por el Sol se encuentran:

- a) *La variación de 11 años*
- b) *El decrecimiento Forbush*
- c) *La variación diurna*

La *variación de 11 años* está sujeta a una variación periódica debida al ciclo solar promedio de 11 años. Existe una anti-correlación entre los rayos cósmicos detectados y el número de manchas solares. Como el número de manchas solares es un indicador de la actividad solar, a mayor número de manchas, mayor actividad y emisiones en el Sol. Las emisiones de la actividad del Sol permean todo el Sistema Solar con líneas de campo magnético, las cuales desvían a los rayos cósmicos que ingresan. Cuando el Sol se encuentra con baja actividad, las emisiones son menos intensas y permiten que el flujo de rayos cósmicos galácticos se incremente. Por lo tanto, a mínimo número de manchas, menor cantidad de emisiones solares y mayor el flujo de rayos cósmicos detectados, con un retardo de aproximadamente 1 o 2 años [22]. En la figura 1.2 se puede ver la anti-correlación entre el número de manchas solares observados por Solar Influences Data Analysis Center (SIDC) y las cuentas del monitor de neutrones de Oulu.

El *decrecimiento Forbush*, llamado así en honor al físico estadounidense Scott E. Forbush, es el decrecimiento repentino de la intensidad de la radiación cósmica hasta en un 10% y ocasionalmente 20 o 30% en el lapso de unas cuantas horas. Una vez que la intensidad ha llegado a un mínimo empieza a recuperarse gradualmente, lo cual puede durar varias horas, días o incluso semanas.

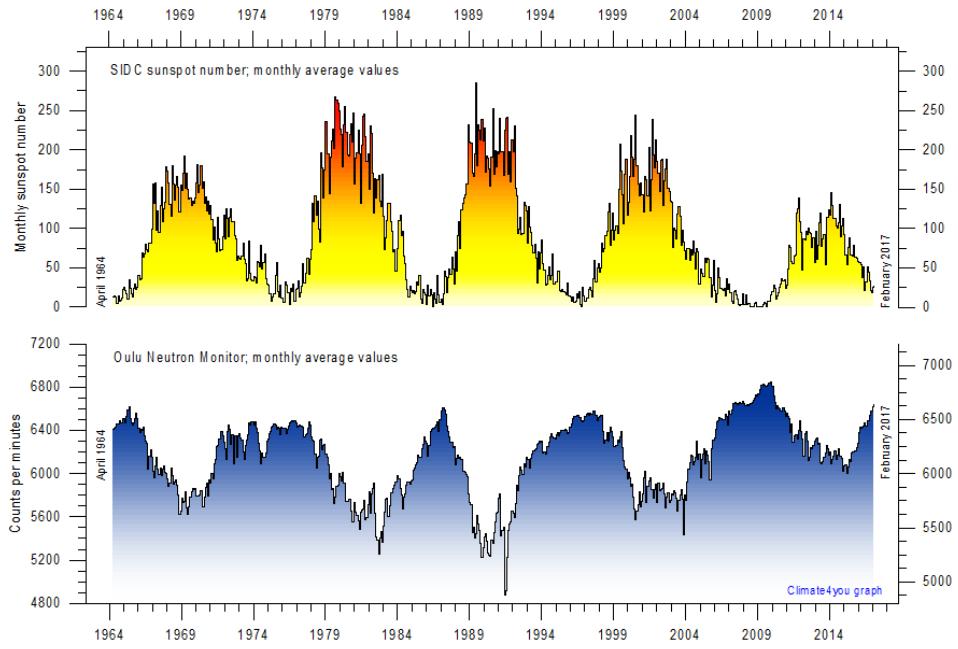


Figura 1.2: Observaciones mensuales del número manchas solares (Solar Influences Data Analysis Center (SIDC)) desde abril de 1964, y medias mensuales de las cuentas del monitor de neutrones de Oulu (Finlandia), ajustados por presión barométrica y eficiencia. Actualización: 1 de marzo de 2017.

Aún se estudian los mecanismos y procesos de generación de los decrecimientos Forbush y se sabe que son un efecto de las líneas de campo magnético asociadas con las emisiones solares. En el momento que ocurre una ráfaga solar o una eyección de masa coronal de una región activa del Sol, la nube de plasma eyectada puede tener una velocidad más alta que el viento solar normal, lo cual produce una discontinuidad brusca conocida como *onda de choque*. La onda de choque forma un “contenedor” magnético con una intensidad de campo relativamente alta. Si la Tierra se encuentra dentro de este contenedor, la radiación cósmica es desviada en función a su energía y no ingresa a la atmósfera; a medida que esta onda se aleja, menos será la influencia sobre las partículas [5, 15]. En la figura 1.3 se muestra un decrecimiento forbush detectado por el monitor de neutrones del Observatorio de Rayos Cósmicos de la Ciudad de México.

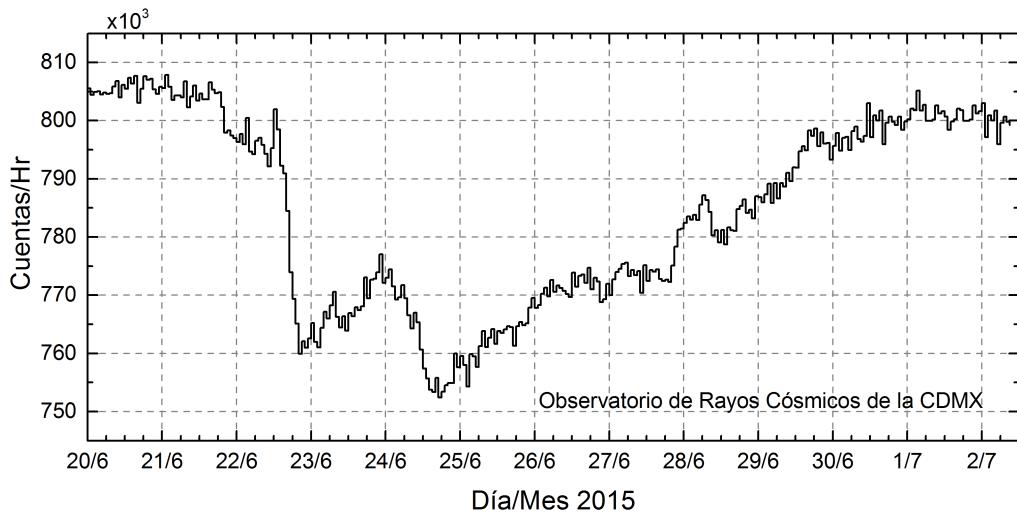


Figura 1.3: Decrecimiento Forbush detectado por el monitor de neutrones de la Ciudad de México el 22 de junio de 2015.

La *variación diurna* es una variación periódica en la intensidad de la radiación cósmica. Tiene una periodicidad de un día solar, que es el tiempo que tarda la Tierra en dar una vuelta sobre su propio eje. El campo magnético gira con el Sol y como los rayos cósmicos giran en espiral alrededor de las líneas de campo, las partículas que se mueven en la misma dirección que la Tierra a lo largo de su órbita tienen un exceso de flujo de aproximadamente 0.7% comparado con la dirección opuesta [5]. Esta variación, al ser observada en una estación terrestre, presenta un máximo y un mínimo de intensidad durante las 24 horas. El máximo ocurre alrededor de la 15 y 18 hrs en tiempo local y un mínimo sobre las 3 de la mañana. Cuando se hacen las correcciones necesarias para tomar en cuenta los efectos del campo geomagnético sobre las partículas de la radiación cósmica, se observa que siempre el máximo de la intensidad se encuentra alrededor de las 18 horas en tiempo local [15].

El TNS y la red mundial de observatorios

Entender el mecanismo de aceleración de partículas es un tema trascendental en la astrofísica, más aún el proceso de aceleración de iones que todavía no se comprende perfectamente. El Sol, siendo el acelerador natural de partículas más cercano, provee de grandes oportunidades para estudiar este proceso [24]. El mejor camino para entender el mecanismo de aceleración de iones en la atmósfera solar es usando datos obtenidos por los detectores de neutrones y rayos gama [12].

Los monitores de neutrones tienen cuentas muy elevadas debido a su alta sensibilidad y no pueden diferenciar la dirección de arriba de las partículas incidentes [23]. Tampoco resuelven la energía de las partículas y no discriminan entre neutrones y protones. Por esta razón, es necesario el uso de un nuevo equipo para resolver los inconvenientes mencionados. El Telescopio de Neutrones Solares (TNS) puede diferenciar entre neutrones y protones, y medir el flujo neutrones solares, su energía y dirección de arriba de las partículas incidentes [3,25].

En este capítulo se muestra la red mundial de Telescopios de Neutrones Solares y se describen las características del TNS instalado en Sierra Negra, así como la estructura de los datos registrados.

2.1. Red mundial de TNS

Para interpretar cualquier medición de intensidad de radiación cósmica que se realice cerca de la superficie de la tierra se requiere tomar en cuenta la presencia del campo magnético terrestre. El campo geomagnético, así como nos protege de los efectos nocivos, también controla la cantidad de rayos cósmicos que llega a la superficie. Por ejemplo, si apuntamos un detector de rayos cósmicos hacia la dirección vertical, se observará que el detector recibe todas las partículas de rigideces altas, como si no estuviera el campo geomagnético. Sin embargo, si seguimos midiendo el flujo de rigideces magnéticas menores veremos que hay una rigidez debajo de la cual ya no se detecta ninguna partícula, a ésta se le conoce como *rigidez umbral*. Además, si desplazamos el detector vertical desde el ecuador hacia los polos veremos que la

rígidez umbral va disminuyendo [15]. De este modo, los rayos cósmicos que llegan a mayores latitudes magnéticas penetran fácilmente a la atmósfera, mientras que los que llegan a bajas latitudes magnéticas requieren más energía para penetrar.

Además de la latitud geomagnética y la energía de partículas, la altura a la cual se encuentra el detector también es muy importante; si se ubica a grandes alturas sobre el nivel del mar, el número de cuentas que registre será mayor debido a que la absorción atmosférica de las partículas secundarias disminuye pues hay menos masa [11].

Si se quiere tener una mayor eficiencia en la detección de neutrones, los telescopios deben ser colocados muy cerca del ecuador para que el tiempo de exposición a la radiación solar sea más uniforme a lo largo del año y que la rigidez umbral requerida para los iones incidentes sea muy alta. También debe estar localizado a la mayor altura posible para reducir la cantidad de materia que puede interaccionar con los neutrones solares, disminuyendo la posible contaminación de partículas energéticas cargadas. Por lo tanto, los sitios ideales para las mediciones es en lo alto de las montañas.

Predecir cuándo habrá un evento de neutrones solares es imposible, por ello se ha instalado una red de telescopios de neutrones solares a diferentes longitudes alrededor del mundo para tener observación del Sol las 24 horas. La red mundial de TNS está construida tomando en cuenta los detalles sobre la localización geográfica para tener una óptima observación. Hasta ahora se han instalado siete detectores de este tipo alrededor del planeta, el más nuevo de ellos es el instalado en México. En la figura 2.1 se puede ver la localización geográfica de cada TNS, así como el país en el que se encuentra instalado.

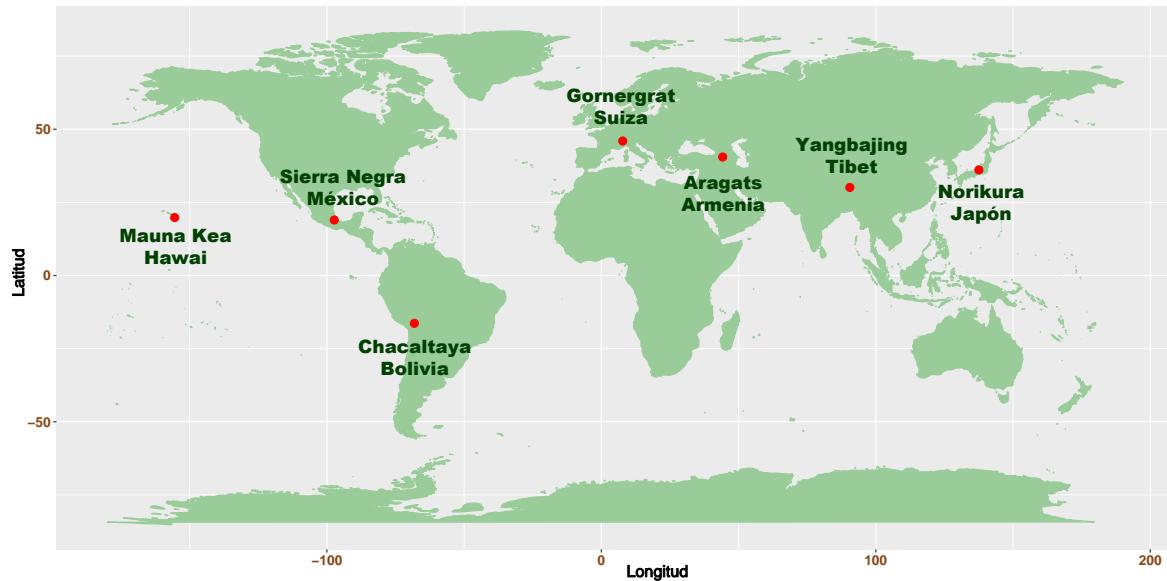


Figura 2.1: Localización geográfica de los TNS alrededor del planeta.

En la tabla 2.1 se muestra las coordenadas y altura de los siete TNS que conforman la red mundial, también la fecha en que se comenzaron a hacer las primeras observaciones. Para ver más características de los TNS se recomienda *Detection efficiency of a new type of solar neutron detector calibrated by an accelerator neutron beam* de H. Tsuchiya [23].

Localización	Altura m.s.n.m	Longitud	Latitud	Inicio de la observación
Gronergrat, Suiza	3135	7.8° E	46.0° N	Enero 1998
Aragats, Armenia	3200	40.5° E	44.2° N	Jun 1997
Yanbajing, Tibet	4300	90.5° E	30.0° N	Sep 1998
Mt. Norikura, Japón	2770	137.5° E	36.1° N	Oct 1990
Mauna Kea, Hawaii	4200	156.3° W	19.8° N	Abr 1997
Sierra Negra, México	4580	97.3° W	19.0° N	Jul 2004
Chacaltaya, Bolivia	5250	68° W	16.2° S	Sep 1992

Tabla 2.1: Datos de la Red Mundial de TNS.

2.2. El TNS en Sierra negra

Sierra Negra es un volcán inactivo localizado en la parte oriental del estado de Puebla, tiene una altura de 4580 m s.n.m., es uno de los picos más altos de la meseta central mexicana. La cima del volcán alberga diversos proyectos astronómicos,¹ uno de ellos es el TNS que forma parte de la red mundial de observatorios.

Debido a su localización geográfica, Sierra Negra fue escogido como el mejor lugar para colocar un TNS en México. Además, las condiciones atmosféricas son lo suficientemente moderadas; así que se preserva el flujo de una buena parte de neutrones solares emitidos en fulguraciones. La instalación se llevó a cabo entre el STELab (Solar Terrestrial Environment Laboratory) de la Universidad de Nagoya, Japón y el Instituto de Geofísica de la UNAM.

Se puede ver en la figura 2.1 que el TNS en Sierra Negra está ubicado, longitudinalmente, entre los detectores de Hawái y Chacaltaya, lo cual ayuda a tener en observación al Sol de

¹El Observatorio Solar Mexicano de Gran Altura (OSOMEGA) de la Universidad Nacional Autónoma de México (<http://naolinco.geofisica.unam.mx/observatorios/osomega/index.html>). También el Gran Telescopio Milimétrico, un proyecto entre el Instituto Nacional de Astrofísica, Óptica y Electrónica (INAOE) y la Universidad de Massachusetts Amherst (<http://www.lmtgtm.org/?lang=es>), entre otros.

manera continua, y la combinación de altura/latitud proporciona un tiempo considerable para tal efecto .

El TNS fue construido y probado en Marzo de 2003 en Sierra Negra, Puebla (97.3° W, 19.0° N) y ha estado operando desde julio de 2004 [9]. La figura 2.2 muestra un esquema del TNS.

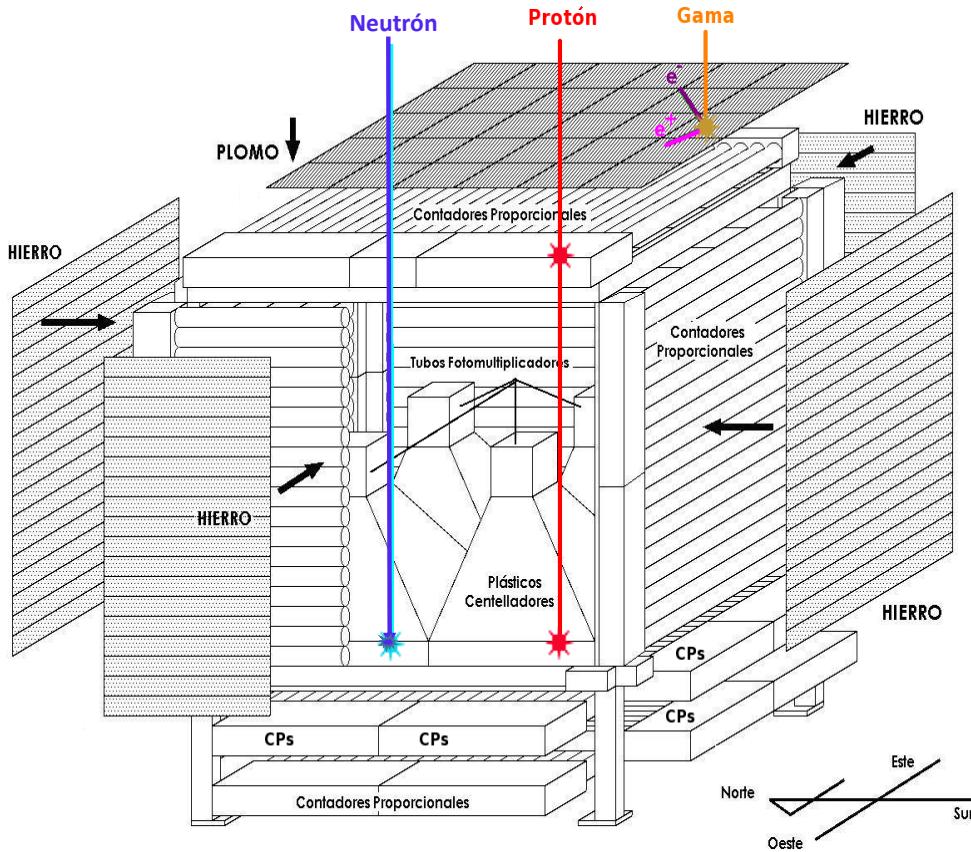


Figura 2.2: Esquema del TNS instalado en la cima del volcán Sierra Negra, Puebla.

El TNS en Sierra Negra consiste de 4 plásticos centelladores (PC) con área de $1m^2$ cada uno y espesor de 30cm; así el área total de detección es de $4m^2$. Los PC están rodeados por arreglos de contadores proporcionales (CP).

Se diferencia entre partículas neutras y partículas cargadas mediante un sistema de anticoincidencias electrónicas entre la señal que disparan los PC y CP. Es decir, las partículas cargadas dejan señal en los plásticos centelladores y en los contadores proporcionales; mientras que las partículas neutras lo hacen sólo en los plásticos centelladores.

La energía depositada por las partículas incidentes se mide en un tubo fotomultiplica-

dor instalado arriba de cada PC, la altura del pulso obtenido por cada fotomultiplicador se discrimina en 4 diferentes canales de deposición de energía; $E \geq 30 \text{ MeV}$ (*S1_with_anti*), $E \geq 60 \text{ MeV}$ (*S2_with_anti*), $E \geq 90 \text{ MeV}$ (*S3_with_anti*), $E \geq 120 \text{ MeV}$ (*S4_with_anti*).

Por encima de los CP se colocó una placa de 0.5 cm de espesor de Plomo; por los lados, los CP se protegieron con placas de Hierro de 0.5 cm de espesor. Estas placas sirven para reducir la radiación de fondo, ya que los fotones muy energéticos pueden dejar una señal similar a la de los neutrones.

Las direcciones de arribo se miden usando cuatro capas de CP alineados ortogonalmente debajo de los PC, dos para determinar la dirección de arribo E-W y dos más para la dirección N-S. Estas direcciones, N-S y E-W están divididas en cinco secciones, con lo cual se tienen 25 canales de dirección. Las direcciones se determinan usando los protones producidos por los neutrones al interaccionar con los PC. Estos protones secundarios se desvian menos de 13.48° de su dirección original con lo que se asegura que la dirección de arribo del neutron puede ser determinada [3].

Hasta aquí se ha visto de qué forma el TNS diferencia entre partículas cargadas y partículas neutras, la manera en que se mide la energía de las partículas, su dirección de arribo y el flujo de partículas neutras. Después de esto, es importante saber si la capacidad de detección del TNS es adecuado, es decir, que las mediciones de neutrones solares son correctas y que no está contaminado por otras partículas.

La eficiencia de detección de los detectores de neutrones solares se calcula mediante una simulación Monte Carlo. El TNS en Sierra Negra tiene una capacidad de detección del 10 %, para neutrones solares muy energéticos, tomando en cuenta los plásticos centelladores. Sin embargo, la eficiencia disminuye si se toma en cuenta los contadores proporcionales y las placas de Hierro y Plomo. En *El Telescopio de Neutrones Solares en Sierra Negra y Aceleración de Iones en la Atmósfera Solar* de Luis Xavier González [3] se pueden ver los detalles de la capacidad de detección del TNS tomando en cuenta los componentes mencionados.

Una vez que el TNS en Sierra Negra estuvo instalado y probado se empezaron a registrar las primeras observaciones en julio de 2004. Despues de 13 años de detección es conveniente estudiar la estabilidad de los datos para distinguir las diferentes variaciones. Para ello es importante conocer la estructura de los datos registrados antes de procesarlos y así saber cómo manipularlos.

2.3. Datos del TNS en Sierra Negra

Es de gran importancia analizar los datos del TNS para tener un resumen de las variaciones que ha presentado su desempeño a lo largo del tiempo. El análisis se realiza para los cuatro canales de deposición de energía que registran partículas neutras, se registra como anti-coincidencia electrónica (*S_with_anti*).

El registro de la detección se guarda en un archivo diario, cada archivo se almacena en la carpeta del mes correspondiente y, a su vez, la carpeta de cada mes en el año correspondiente. El nombre con el que se guarda el archivo es la fecha del día con extensión **sn1** si no hubo ninguna suspensión en el registro, de lo contrario se generan i archivos con extensión snj (donde $j = 1, \dots, i$) para $i - 1$ interrupciones. La ventaja de este tipo de extensión es que los datos se pueden visualizar en cualquier editor de textos y datos en ascii.

La razón de conteo desde enero de 2004 al 6 de marzo de 2006 fue de 1 minuto, y a partir del 8 de Marzo de 2006 los registros se hicieron cada 10 segundos. Excepto del 5 de marzo de 2008 al 24 de abril de 2008 que fue de 3 segundos.

Debido a fallas electrónicas en el equipo periférico y de suministro de energía eléctrica hay muchos archivos en los que se suspendió el registro por algunas horas, otros en los que se registraron datos de 0 en todo el día y días en los que no se registró la detección del TNS. Para esto se tomó nota de los archivos completos (registro de las 24 horas), de los que se suspendieron y la hora en que sucedió, para los años 2004 hasta 2016.

Durante todo el año del 2007 el TNS estuvo desconectado. Así, desde 2004 a 2016 no hay registros de 1205 días. La figura 2.3 muestra el número de días registrados por cada mes y año. Con esta gráfica se puede observar fácilmente en qué meses hay pocos registros.

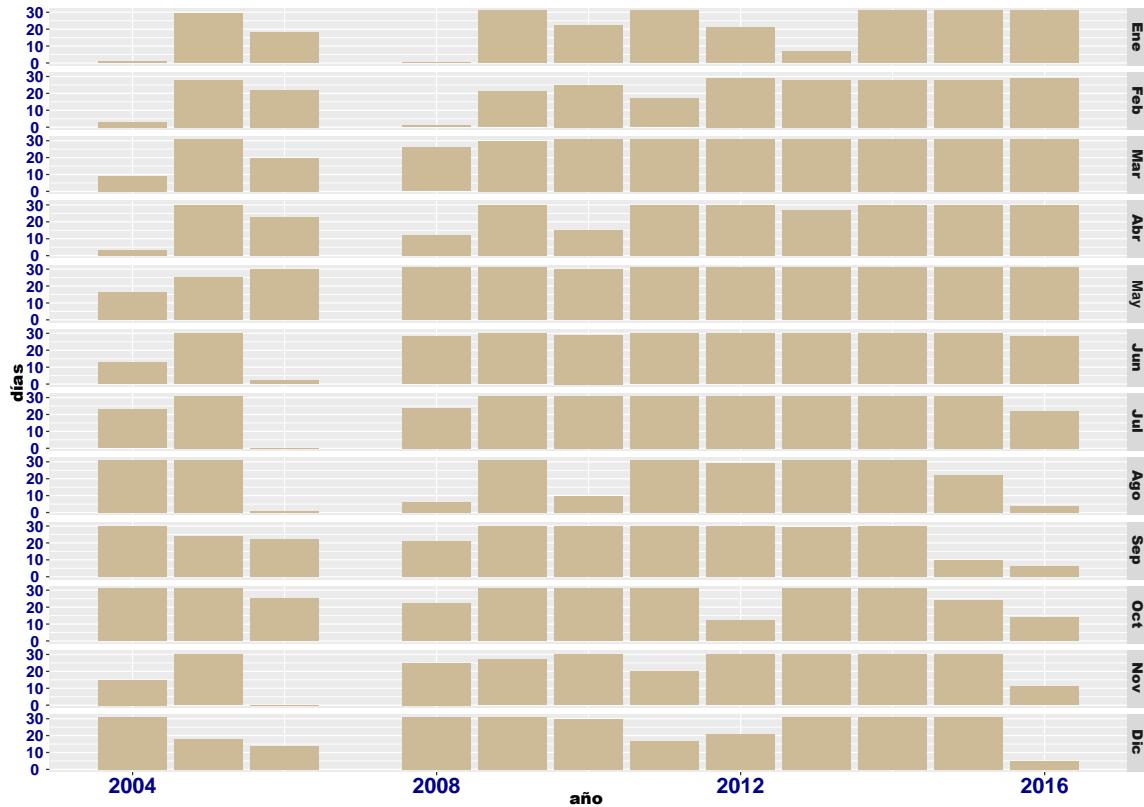


Figura 2.3: Gráfica que muestra el número de días registrados por mes y año.

Es importante destacar que la gráfica sólo muestra la cantidad de días en que hubo registros, lo cual no quiere decir que todos los datos registrados sean “buenos”. Por ejemplo, en abril del 2010 hubo varios cortes de energía que duraron más de un día. Así varios archivos no tienen datos, los datos son constantes o los archivos están interrumpidos porque el soporte de energía era de 18 horas y el corte respectivo fue mayor.

También en septiembre de 2010, varios días tienen datos con ceros, debido a cortes en el suministro de energía eléctrica. Este problema persistió hasta principios de 2011. En el *capítulo 3* se mostrará el porcentaje de datos confiables.

2.4. Estructura de archivos registrados

Antes de pasar al siguiente capítulo para mostrar los códigos usados al importar los datos, es necesario conocer cómo están estructurados. Se han manejado varias razones de conteo: de 3 segundos, 10 segundos y un minuto.

En la Tabla 2.2 se muestran las primeras y últimas filas del archivo mx040720.sn1, que

contiene datos del 20 de julio de 2004. La primera columna denota el año en que se está llevando a cabo el registro, en este caso 2004; la segunda columna contiene el número del mes y el día, es decir, 720 indica que es el día 20 del mes 7 (julio). En la columna 3 se encuentra el tiempo, empieza con el minuto 0, sigue el minuto 1, 2, 3 , 4 hasta llegar al minuto 59 y luego 100 que denota la hora 1 con cero minutos. Se sigue con 101,102 hasta llegar a 159, luego 200, que es la hora 2 con cero minutos. De esta manera, llega a la fila 1440 a la hora 23 con 59 minutos y termina con el minuto 0 del siguiente día.

La cuarta columna no es ningún dato relevante. De la columna 5 a la 8 son los datos de partículas neutras (canales S1_A, ..., S4_A), de la columna 9 a la 12 son datos de partículas cargadas desde S1 hasta S4. Luego siguen otras 378 columnas con registro de otros datos que no se usarán para este trabajo.

	1	2	3	4	5	6	7	8	9	10	11	...
1	2004	720	0	100	8761	3407	1196	541	27371	12855	5036	...
2	2004	720	1	0	8602	3355	1187	539	27619	13003	5083	...
3	2004	720	2	0	8705	3402	1261	580	27804	13205	5194	...
4	2004	720	3	0	8662	3400	1235	565	27645	12987	5051	...
5	2004	720	4	0	8712	3323	1178	551	27702	12876	4994	...
6	2004	720	5	0	8642	3271	1210	553	27627	12857	5107	...
:	:	:	:	:	:	:	:	:	:	:	:	:
1436	2004	720	2355	0	8577	3316	1204	569	27647	13057	5104	...
1437	2004	720	2356	0	8652	3358	1233	561	27647	12928	5083	...
1438	2004	720	2357	0	8665	3400	1231	545	27596	12999	5039	...
1439	2004	720	2358	0	8732	3399	1226	548	27675	12983	5060	...
1440	2004	720	2359	0	8681	3355	1152	557	27789	12983	5009	...
1441	2004	721	0	0	8651	3377	1252	558	27848	13032	5098	...

Tabla 2.2: Datos del 20 de julio de 2004, razón de conteo de 1 minuto.

Ahora véase la Tabla 2.3, con datos del 1 de junio de 2008, esta es la estructura de los archivos registrados cada 10 segundos. Las primeras cinco filas contienen información acerca del registro de ese día, como *INTERVAL* que indica el intervalo de tiempo entre cada observación, *DATE* contiene la fecha en el formato año, mes y día (080601, 1 de junio de 2008). *TIME* es la hora o tiempo en que empieza el registro, en este caso 000000 indica que empieza a las 0 horas, 0 minutos con 0 segundos. Mientras que la última fila (fila 8646) dice que el archivo termina con 8640 observaciones.

Como se puede ver, las primeras cinco filas son innecesarias, ya que la columna 2 contiene la fecha en el mismo formato para cada observación, así como la columna 3 guarda la hora, minuto y segundo. De la columna 4 a la 7 son datos de partículas neutras (canales S1_A, ..., S4_A), mientras que la columna 8 a la 11 guarda datos de partículas cargadas (S1, ..., S4). En total el archivo contiene 52 columnas.

	1	2	3	4	5	6	7	8	9	...
1	NCHANNEL	48								
2	INTERVAL	10								
3	VERSION	3.00								
4	DATE	080601								
5	TIME	000000								
6	SG	080601	000010	12400	4120	1108	108	26692	11292	...
7	SG	080601	000020	12408	4100	1036	12	26656	11372	...
8	SG	080601	000030	12364	4124	1104	104	26636	11344	...
9	SG	080601	000040	12292	4108	1024	124	26664	10292	...
10	SG	080601	000050	12348	4216	1148	60	26632	11364	...
11	SG	080601	000100	12412	4188	1028	0	26740	10260	...
:	:	:	:	:	:	:	:	:	:	:
8641	SG	080601	235920	12408	4108	1136	4	26704	10316	...
8642	SG	080601	235930	12412	4096	1120	100	26656	10320	...
8643	SG	080601	235940	12364	4140	1144	8	26732	11352	...
8644	SG	080601	235950	12400	4128	1072	120	26648	11356	...
8645	SG	080602	000000	12316	4100	1060	96	26740	10284	...
8646	END		8640							

Tabla 2.3: Datos del 1 de junio de 2008, razón de conteo 10 segundos.

Nótese en la Tabla 2.4 que la estructura para archivos con razón de conteo 3 segundos es igual para los archivos con datos de registro cada 10 segundos. Lo único en que difieren es en la cantidad de filas u observaciones.

	1	2	3	4	5	6	7	8	9	...
1	NCHANNEL	48								
2	INTERVAL	2								
3	VERSION	3.00								
4	DATE	080310								
5	TIME	000000								
6	SG	080310	000003	4120	1104	88	80	7184	3164	...
7	SG	080310	000006	4172	1148	112	84	8248	3144	...
8	SG	080310	000009	4192	1140	120	104	8196	3112	...
9	SG	080310	000012	4148	1132	112	76	7224	3176	...
10	SG	080310	000015	4196	1104	108	116	8240	3080	...
:	:	:	:	:	:	:	:	:	:	...
28802	SG	080310	235951	3196	1060	12	0	7288	3076	...
28803	SG	080310	235954	4124	1060	116	108	7212	3192	...
28804	SG	080310	235957	4116	1120	0	100	8288	3096	...
28805	SG	080311	000000	4200	1124	124	100	7256	3152	...
28806	END	28800								

Tabla 2.4: Datos del 10 de marzo de 2008, razón de conteo 3 segundos.

En el *capítulo 3* se explica el código para realizar esta lectura de datos en R.

Ahora que se conoce la estructura de los datos, el objetivo es prepararlos para el posterior análisis. Para un manejo más eficiente, es necesario tener una sola columna que contenga la fecha y hora, así como una columna con el respectivo nombre de la variable para cada uno de los canales de deposición de energía. Al final se tendrá un “data frame” de 5 columnas o variables con observaciones por año. A la primera columna se le llamará “date”, representará la fecha y hora. De la segunda a la quinta columna se les llamará S1_A, S2_A, S3_A, S4_A respectivamente para cada canal de energía.

Como el registro se lleva a cabo en tiempo local, hay que tener especial cuidado al convertir los datos de fechas y tiempo a clase “date/time” en el ambiente R (Ver Capítulo 3). Para entender el procedimiento de limpieza y ordenamiento de datos, en el capítulo 3 se habla de manera resumida la sintaxis que se maneja en R para después mostrar y explicar el código usado al procesar, ordenar y realizar el análisis de datos.

Capítulo 3

R y desarrollo de códigos

R es un lenguaje de programación para análisis estadístico y gráfico creado por Ross Ihaka y Robert Gentleman [18], se distribuye gratuitamente bajo los términos de *GNU General Public Licence*¹ y puede verse como una implementación alternativa del lenguaje S, desarrollado en AT&T Bell Laboratories. S está disponible como el programa S-PLUS comercializado por Insightful².

R tiene una naturaleza doble de programa y lenguaje de programación. Como entorno de programación se desarrolla mediante paquetes (*packages*) o bibliotecas que complementan el lenguaje con nuevos desarrollos para distintas áreas de análisis estadístico y gráfico de datos. Algunos se encuentran en el *sistema base* pero muchos otros están disponibles en el sitio de internet *Comprehensive R Archive Network* (CRAN) que extienden la funcionalidad. Básicamente se trata de una ventana de trabajo (consola), se pueden iniciar distintas sesiones de trabajo que podemos guardar para retomar con posterioridad en el punto que fue dejado. Los scripts consisten en una serie de instrucciones modificadas por comandos y variables ejecutadas sobre un conjunto de datos [16].

Los archivos necesarios para instalar R, ya sea desde las fuentes o binarios pre-compilados, se distribuyen desde CRAN³ junto con las instrucciones de instalación. R está disponible en varias formas: el código fuente escrito principalmente en C (y algunas rutinas en Fortran), esencialmente para máquinas Unix y Linux, o como archivos binarios pre-compilados para Windows, Linux (Debian, Mandrake, RedHat, SuSe), Macintosh y Alpha Unix.

En resumen, R proporciona un entorno de trabajo especialmente diseñado para el análisis estadístico de datos. Sus principales características son las siguientes: R proporciona un lenguaje de programación propio, basado en el lenguaje S, que a su vez tiene muchos elementos del lenguaje C; sin embargo, la semántica es muy distinta a la de este último. Esto es porque R permite ejecuciones de comandos en línea (compilación y ejecución unidos en un mismo paso).

¹Ver <http://www.gnu.org/licenses/gpl.html>

²Para más información ver <http://www.insightful.com/products/splus/default.html>

³En la web <https://cran.r-project.org/>

Como cualquier lenguaje de programación, R tiene sus ventajas y desventajas:

Ventajas

- R es un software libre, se refiere a la libertad de los usuarios para ejecutar, copiar, distribuir, estudiar, cambiar y mejorar el software.
- Es multiplataforma.
- Es de código abierto.
- Actualización constante.
- R es una plataforma estadística, se trata de un lenguaje creado específicamente para la visualización y exploración de datos, así que cuenta con muchas funciones de herramienta estadística.
- Los gráficos son de calidad.

Desventajas

- Vasta documentación, esto dificulta encontrar información específica rápidamente.
- Los mensajes de error no son tan claros.
- Lenguaje de programación en línea de comando, es una desventaja para los que no tienen nociones de programación, aunque ayuda a entender mejor la base de la estadística.

Debido a las ventajas y herramientas de cómputo estadístico que ofrece R para la manipulación y gráfico de datos, se escogió este software para trabajar con los 13 años de detección del TNS para los cuatro canales que registran partículas neutras. En este capítulo se describirá el método utilizado, así como los paquetes usados y códigos principales para el procesamiento y manipulación de los datos. Para entrar en contexto con el lenguaje de R se da una breve introducción a la sintaxis.

3.1. Entorno de programación R

R es un lenguaje Orientado a Objetos, además es un lenguaje interpretado (como Java) y no compilado (como C, C++, Fortran, Pascal), lo cual significa que los comandos escritos en el teclado son ejecutados directamente sin necesidad de construir ejecutables. Todas las acciones en R se realizan con objetos que son guardados en la memoria activa del ordenador, no se usan archivos temporales [16]. La lectura y escritura de archivos solo se realiza para la entrada y salida de datos y resultados (gráficas, etc). Los resultados se pueden visualizar directamente en la pantalla, guardar en un objeto o escribir directamente en el disco (particularmente para gráficos). Debido a que los resultados mismos son objetos, pueden ser considerados como datos y analizarlos como tal. Archivos que contengan datos pueden ser leídos directamente desde el

disco local o en un servidor remoto a través de la red [16].

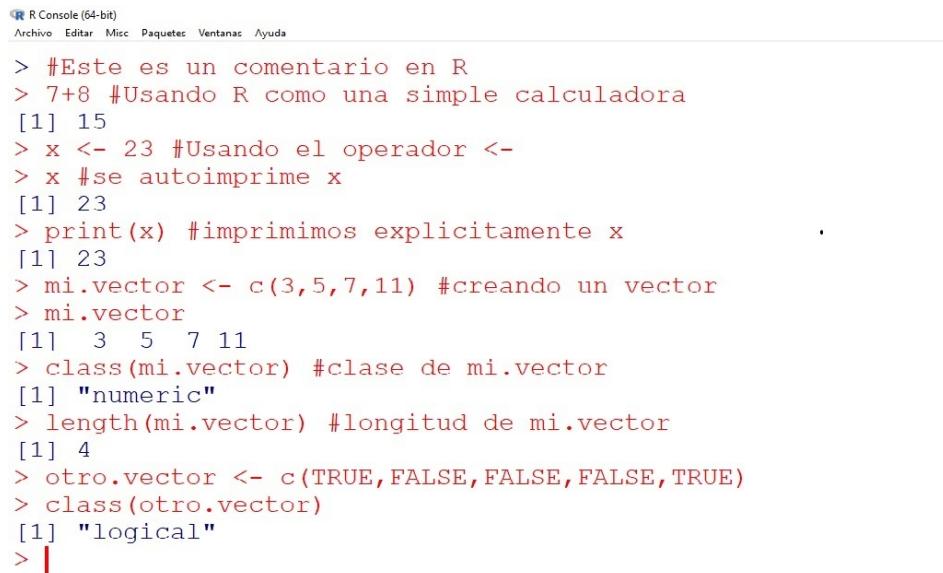
Hay diferentes tipos o clases de datos en R:

- *character*: cadenas de caracteres.
- *numeric*: números reales.
- *integer*: números enteros.
- *complex*: números complejos.
- *logical*: valor lógico falso (*FALSE*) o verdadero (*TRUE*).

Casi todo en R es un objeto, como son las operaciones (aritméticas o lógicas), las funciones; que constan de una lista de argumentos que se usan para realizar ciertas acciones y regresar un resultado, entre otros. El objeto más básico que puede contener alguno de los tipos de datos anteriores es el *vector*, éste puede contener cero o más elementos que deben ser de la misma clase. Una manera de crear vectores es a partir de los elementos individuales que compondrán el vector [20]. Para esto se usa la función *c()* como se muestra en la Figura 3.1.

Una característica que pueden tener los objetos son los *atributos*, éstos pueden ser de diferentes tipos como: nombres, dimensiones, clase, longitud, entre otros. Por ejemplo, las *matrices* son un tipo de vector especial con un atributo de dimensión que indica el número de renglones y de columnas.

El símbolo `<-` es el símbolo de asignación. Para ejecutar una expresión en la consola se presiona “Enter”.



```
R Console (64-bit)
Archivo Editar Misc Paquetes Ventanas Ayuda

> #Este es un comentario en R
> 7+8 #Usando R como una simple calculadora
[1] 15
> x <- 23 #Usando el operador <-
> x #se autoimprime x
[1] 23
> print(x) #imprimimos explicitamente x
[1] 23
> mi.vector <- c(3,5,7,11) #creando un vector
> mi.vector
[1] 3 5 7 11
> class(mi.vector) #clase de mi.vector
[1] "numeric"
> length(mi.vector) #longitud de mi.vector
[1] 4
> otro.vector <- c(TRUE, FALSE, FALSE, FALSE, TRUE)
> class(otro.vector)
[1] "logical"
> |
```

Figura 3.1: Entrada de datos en R. El [1] que aparece a la izquierda de la salida de la expresión indica que el elemento que se encuentra a la derecha tiene índice 1.

Otros tipos de objetos son las *listas* y los *dataframes*. Una lista, de la clase *list*, es una clase de datos que puede contener elementos de diferentes clases y de distinta longitud o dimensión. Por otro lado, un dataframe es un tipo de dato en R que sirve para guardar datos tabulares donde cada columna representa una variable, las columnas no necesariamente son de la misma clase pero deben tener la misma longitud. Para generar un dataframe se usa *data.frame()*, *read.table()*, *read.csv()*. Ver ejemplos en la Figura 3.2.



```
R Console (64-bit)
Archivo Editar Misc Paquetes Ventanas Ayuda

> #creando una lista de 2 elementos
> mi.lista <- list(mes=c("Enero", "Marzo"), dia=c(12, 15))
> mi.lista
$mes
[1] "Enero" "Marzo"

$dia
[1] 12 15

> class(mi.lista)
[1] "list"
> #creando un dataframe de 2 columnas
> datos <- data.frame(var1=c(12,17,21), var2=c(16,9,18))
> datos
  var1 var2
1    12   16
2    17    9
3    21   18
> class(datos)
[1] "data.frame"
> |
```

Figura 3.2: Creando listas y dataframes.

Para el manejo de fechas y tiempo hay clases especiales dentro del sistema. Este tipo de objetos permiten llevar a cabo operaciones numéricas y estadísticas. Hay un tipo de datos exclusivamente para fechas usando la función *as.Date()*. Además, hay 2 tipos de datos para tiempo: *POSIXct* y *POSIXlt*. La diferencia entre estos dos es la manera en que almacenan los datos internamente en R, el primero almacena el dato como el número de segundos transcurridos desde el 1 de enero de 1970, mientras que *POSIXlt* almacena los datos de la fecha y tiempo en forma de lista.

Para manipular la información de fecha y tiempo de los datos del TNS se usará la biblioteca *lubridate* porque es más eficiente al hacer la conversión en grandes cantidades de datos que el paquete base de R. Más adelante se explica detalladamente este procedimiento.

En el capítulo 2 se vio cuál es la estructura de los datos, el objetivo de este capítulo es mostrar y explicar a grandes rasgos el código empleado para obtener un dataframe ordenado y el método para hacer la limpieza de datos.

3.2. Limpieza y ordenamiento de datos

La limpieza de datos es una parte esencial del análisis estadístico. De hecho, en la práctica lleva aún más tiempo que el análisis en sí. La limpieza de datos o *data cleaning* es el proceso de transformación de datos crudos en datos consistentes que puedan ser analizados. Si bien el orden de las variables y las observaciones no afectan al análisis, un buen orden hace que sea más fácil que un analista o una computadora extraigan las variables necesarias porque proporciona una forma estándar de estructurar un conjunto de datos.

De acuerdo al artículo *Tidy Data* de Hadley Wickham [28] los datos ordenados deben cumplir con lo siguiente:

1. Cada variable forma una columna.
2. Cada observación forma una fila.
3. Cada tipo de unidad observacional forma una tabla.

En esta sección describiré como ordené los datos del TNS de acuerdo a los principios mencionados anteriormente y extrayendo sólo las variables que nos interesan (datos que registran partículas neutras y la fecha y tiempo).

El lenguaje R puede leer datos guardados como archivos de texto plano, del tipo *.csv*, *.txt*, *.dat*. También lee archivos en otros formatos (Excel, SAS, SPSS) pero las funciones necesarias no están incluidas en el paquete *base*. Para leer y escribir archivos, R utiliza el directorio de trabajo. Para conocer el directorio se utiliza el comando *getwd()*¹. Para cambiar el directorio de trabajo se utiliza la función *setwd()* utilizando como argumento la dirección (path).

En el capítulo anterior se comentó que los datos del TNS se guardan en archivos con extensión *.sn1*, estos son archivos de texto en formato ASCII que se pueden leer con las funciones *read.table()*, *read.csv()*, *read.csv2()*; las cuales regresan un objeto de tipo *data frame*. Como no es necesario tener el conjunto completo de datos (13 años) todo el tiempo en una sesión de R, se procesó la información por año.

Para leer y ordenar los datos del TNS se usaron dos códigos distintos, uno para los datos registrados por minuto y otro para los que se registraron cada 3 y 10 segundos. Estos últimos aún cuando no tengan la misma razón de conteo, los archivos sí tienen la misma estructura por lo que no hay necesidad de tratarlos por separado, ya que en R se pueden hacer operaciones con fechas y tiempos para ponerlos en una misma razón de conteo cuando sea necesario.

A continuación se explican los códigos mencionados anteriormente, el primero es un ejemplo para datos de 2005 y el segundo es para el año 2008. Para los demás años se procede de la misma forma que en alguno de los dos casos.

¹get working directory

Procesamiento de datos para 2005:

Primero se cambia el directorio de trabajo a la carpeta donde se encuentran los archivos de 2005. En la línea 2 se obtienen los nombres de estos archivos y se guarda en la variable “archivos”. En la línea 3 se crea la función *leer1*, ésta recibe un parámetro que se usa en la función *read.table()* y solo va a leer las columnas del 1 al 3 y luego del 5 al 12. En la línea 4, con la función *lapply*, se aplica la función *leer1* con cada una de las entradas del vector *archivos*, el resultado es una lista con elementos que son dataframes, se guarda en *dat05*. Por último, en la línea 5 se extrae cada entrada de la lista *dat05* y con *rbind* cada dataframe se “pega” uno debajo de otro. El resultado es un dataframe que contiene todos los datos de 2005 con las columnas del 1 al 3 y del 5 al 12.

```

1 > setwd("./Data_TNS/2005")
2 > archivos <- list.files()
3 > leer1 <- function(file){read.table(file)[c(1:3,5:12)]}
4 > dat05 <- lapply(archivos,leer1)
5 > dat05 <- do.call(rbind,dat05)

```

Ya se tienen los datos del año 2005 en un dataframe, falta poner en una sola columna la información de fecha y tiempo para cada observación en el formato año-mes-día hora:minuto:segundo, se hace en el *Paso 2* y *Paso 3*. La primera línea calcula la cantidad de “ceros” que le falta a cada elemento de la columna 3 para que sean 4 dígitos, y devuelve esa cantidad de ceros en un vector *temp*. La segunda línea pega el vector *temp* con la columna 3 de *dat05* y otros dos ceros al final, esta operación es elemento a elemento y se “reescribe” en la columna 3. La tercera línea muestra las primeras 4 filas del resultado.

Paso 2:

```

> temp <- mapply(function(x, y) paste0(rep(x, y),
collapse = ""), 0, 4 - nchar(dat05$V3))
> dat05$V3 <- paste0(temp, dat05$V3,00)
> head(dat05,4)

      V1   V2   V3     V5   V6   V7   V8     V9   V10  V11   V12
1 2005 101 000000  9147 3365 1217 561 27783 12932 5109 2476
2 2005 101 000100  9099 3458 1215 574 27809 12872 5039 2446
3 2005 101 000200  8979 3415 1200 526 27854 13016 5014 2415
4 2005 101 000300  8981 3434 1194 573 27917 13013 5079 2484

```

Para el *Paso 3* se instaló previamente la biblioteca *dplyr* (se explica en la siguiente sección), se carga a la sesión de R en la primera línea. La segunda línea es una nueva gramática

del paquete dplyr, lo primero que hace es tomar el dataframe dat05, con mutate le agrega una nueva columna llamada *date* que es el resultado de pegar las columnas V1,V2 y V3. Posteriormente, con *select* extrae solo las variables de interés, date y datos de los ocho canales. Luego, con *rename* le pone nombres a las variables. Por último, se muestran las primeras cuatro líneas del resultado y en la última línea guarda el dataframe en una ventana de datos de R, en dat05.RData.

Paso 3:

```
> library(dplyr)
> dat05 <- dat05 %>% mutate(date = paste(V1,0,V2," ",V3)) %>%
  select(date,V5:V12) %>%
  rename(S1_A=V5,S2_A=V6,S3_A=V7,S4_A=V8) %>%
  rename(S1=V9,S2=V10,S3=V11,S4=V12)
> head(dat05,4)
      date   S1_A   S2_A   S3_A   S4_A   S1     S2     S3     S4
1 20050101 000000  9147 3365 1217 561 27783 12932 5109 2476
2 20050101 000100  9099 3458 1215 574 27809 12872 5039 2446
3 20050101 000200  8979 3415 1200 526 27854 13016 5014 2415
4 20050101 000300  8981 3434 1194 573 27917 13013 5079 2484
> save(dat05,file="dat05.RData")
```

Ahora véase el ejemplo para el año 2008, el primer paso es igual que el ejemplo anterior, solo difieren en la función *leer2* porque los archivos tienen otra estructura, en este caso se leerán las columnas del 2 al 11.

Ejemplo para datos de 2008.

```
1 > setwd("./Data_TNS/2008")
2 > archivos <- list.files()
3 > leer2 <- function(file){read.csv2(file,header=F,skip=5,sep=" ",
4   colClasses=c(rep("character",3),rep("integer",49)))[2:11]}
5 > dat08 <- lapply(archivos,leer2)
6 > dat08 <- do.call(rbind,dat08)
```

El siguiente paso es similar al Paso 3 del ejemplo anterior. Se toma el dataframe dat08, con *mutate* se le agrega la columna date y con *select* se extraen las columnas date y datos de

los ocho canales. Al final se guarda el dataframe en la ventana de datos *dat08.RData*.

Paso 2:

```
> library(dplyr)
> data08 <- dat08 %>% mutate(date = paste(V2,V3)) %>%
  select(date,V4:V11) %>%
  rename(S1_A=V4,S2_A=V5,S3_A=V6,S4_A=V7) %>%
  rename(S1=V8,S2=V9,S3=V10,S4=V11)
> save(dat08,file="dat08.RData")
```

Como se ha mencionado anteriormente, hay varios paquetes que se han desarrollado para complementar el lenguaje de R. Si bien muchos paquetes desarrollan nuevos algoritmos, hay otros que realizan acciones que se pueden llevar a cabo con el paquete base. La ventaja de estas bibliotecas es que reducen el tiempo de ejecución cuando se trabaja con grandes cantidades de datos y el lenguaje es más intuitivo, por ejemplo *lubridate* y *dplyr*.

Hasta este punto se tienen los datos ordenados del TNS para cada año y todos tienen la misma estructura, pero falta “decirle” a R que la columna *date* de cada dataframe es una columna de fechas y tiempo. Para convertir esta columna en clase “date/time” se usará la biblioteca *lubridate*, el código se muestra en la siguiente sección así como una breve descripción de los paquetes que permitieron trabajar los datos del TNS de manera eficiente.

3.3. Paquetes usados de R

Para usar los paquetes que no están incluidos en el *paquete base* es necesario instalarlos. Para ello se usa la función *install.packages()* con argumento entre comillas la biblioteca que se desea instalar. Después con la función *library()* se carga la biblioteca al área de trabajo.

Ejemplo:

```
>install.packages("dplyr")
>library(dplyr)
```

A continuación se describen algunos paquetes en el orden en que fueron utilizados para llevar a cabo este trabajo.

3.3.1. dplyr

dplyr es una nueva gramática de manipulación de datos, una herramienta rápida, flexible y consistente para trabajar con datos como objetos. Entre las funciones más importantes se encuentran *mutate*, *select*, *rename*, *filter*, *arrange* y *group_by* [29]. Estas funciones fueron muy útiles al procesar los datos, como se vio en los ejemplos anteriores, ya que reduce el tiempo de ejecución comparado con las funciones que realizan la misma acción en el paquete base. Además el lenguaje es más claro y reduce líneas de código.

3.3.2. lubridate

Lubridate es un paquete de R, creado por Garrett Grolemund y Hadley Wickham, que hace mucho más fácil el trabajo con datos de fechas y horas. Identifica el orden en que aparece el año, mes y día en los datos. Además de que facilita el trabajo con zonas horarias, por defecto lubridate pone la fecha en tiempo universal (TU). En el link se pueden ver ejemplos de las herramientas que ofrece este paquete: <https://cran.r-project.org/web/packages/lubridate/vignettes/lubridate.html> o bien en *Dates and times made easy with lubridate* de Garrett Grolemund y Hadley Wickham [6].

Se usará esta biblioteca para convertir a clase POSIXct la columna *date* de cada tabla de datos del TNS. Como estas tablas ya tienen la misma estructura el siguiente código funciona para todas. Lo primero que se hace es localizar los índices de los datos que se registraron en horario de verano y los que no, de acuerdo a estos índices se crea un vector que contiene CDT (Central Daylight Time) si el dato corresponde al horario de verano y CST (Central Standard Time) si no (esta parte se trabajó de manera interactiva localizando las fechas del cambio de horario).

En el siguiente ejemplo primero se carga la biblioteca lubridate. En la segunda línea, la función *merge* junta dat05 con el dataframe que tiene por columnas TZONE y OFFSET. Si la observación tiene asignado el carácter “CDT”, en la columna OFFSET del dataframe dat05 se le asigna -5 y se le asigna -6 si tiene “CST”.

Ejemplo para *dat05*

```

1 > library("lubridate")
2 > data05 <- merge(dat05, data.frame(TZONE = c("CDT", "CST"),
3                               OFFSET = c(-5, -6)))
4 > data_1$DATE_UTC <- ymd_hm(data_1$date,
5                               tz = "UTC") - dhours(data_1$OFFSET)
6 > data_1$DATE_LOCAL <- with_tz(data_1$DATE_UTC, "America/Mexico_City")
7 > data_1<- data_1[order(data_1$DATE_UTC), ]
```

En la cuarta línea se convierte la columna *date* a clase POSIXct pero en Tiempo Universal (UTC), por lo que se le suma la cantidad de horas que faltan, estas horas se encuentran en la columna OFFSET. En la quinta línea, con la función *with_tz* la columna DATE_UTC se cambia a la zona horaria de “America/Mexico_City”, a tiempo local, y finalmente se ordenan los datos de acuerdo a la hora.

Es necesario convertir las fechas y horas primero en tiempo universal y después a tiempo local, ya que si se hace lo último directamente se producen NA’s en las horas repetidas, lo cual hace que la serie sea discontinua cuando si se tiene información de esos datos.

Una vez que se tuvieron los datos ordenados se procedió con la limpieza. Esta parte se trabajó de manera interactiva, explorando cómo se distribuyen los datos y detectando los valores atípicos.

Se observó que al procesar la información, las variables de partículas cargadas no se eliminaron. Esto debido a que se encontraron errores en el registro, había observaciones para las cuales los datos de partículas neutras eran exactamente igual a los datos de partículas cargadas, es decir, el dato del canal S1_A igual al S1, el S2_A igual al S2 y así hasta el S4_A igual al S4. Además, estos datos estaban en el rango de las variables de partículas cargadas, así que se eliminaron esos índices.

Los ceros y datos que se repetían durante varios minutos en todos los canales se sustituyeron por NA (Not Available), que es la forma en que se denotan los valores faltantes en R, ya que era evidente que se trataba de errores electrónicos.

También era fácil ver que se trataba de errores aquellos datos que incrementaban y disminuían repentinamente hasta un 20% con respecto al dato anterior y posterior. Al hablar de estabilidad de datos, lo que se busca es que la variación sea pequeña. En caso de encontrar alguna variación significativa investigar qué pudo haber afectado la toma de datos. Cuando ya no eran tan evidentes los errores se tenía que graficar la serie para ver si había variaciones importantes. Para esto, fue de mucha utilidad el siguiente paquete.

3.3.3. Openair

La biblioteca Openair está diseñada para hacer análisis de datos de calidad del aire, sin embargo ofrece herramientas muy útiles para hacer análisis de variación en series de tiempo, sobretodo la función para agregar datos en distintos intervalos de tiempo. Es por ello que se escogió este paquete para trabajar los datos del TNS. Se puede ver más acerca del paquete en *openair — An R package for air quality data analysis* [1] y en *Lenguaje R aplicado al análisis de Calidad de Aire* [17].

Para hacer las gráficas de variación del capítulo 4 se creó la función *variacion* que recibe dos argumentos, el primero es el dataframe y el segundo es el canal que se desea graficar. Luego estos parámetros se usan en la función *timeVariation* de openair.

```
> library("openair")
> variacion <- function(datos,canal){
  timeVariation(datos, pollutant=canal,
  main="Variación temporal del flujo de partículas neutras",
  xlab=c("Flujo horario durante la semana","Flujo horario",
  "Flujo mensual","Variación por días de la semana"))
}
```

Así, para hacer una gráfica de variación de los datos de 2005 del canal S1_A se hace de la siguiente forma.

```
> variacion("dat05","S1\_A")
```

Hasta este punto se ha cumplido el objetivo de tener datos ordenados y que la fecha y hora de éstos sea reconocida como tal en R, además se limpian los datos de partículas neutras que eran igual a los datos de partículas cargadas. Luego se extrajeron solo la columna date y datos de los cuatro canales de interés. A partir de aquí se empezaron a localizar los valores atípicos y ver si se debían a errores electrónicos. Sin embargo, había casos en los que el intervalo de errores era demasiado grande y por eso los datos “correctos.” eran los que se mostraban fuera de rango. Como en el caso de los datos de 2008, hasta que se graficaron los datos se observó su comportamiento. Véase el siguiente capítulo para ver las gráficas de variación por año.

Análisis de estabilidad

4.1. Introducción

A lo largo de 13 años el TNS ha estado registrando información acerca de la radiación cósmica. Debido a su importancia, como ya se ha mencionado anteriormente, este trabajo es para analizar los datos de partículas neutras. Lo que se pretende es diferenciar las causas de las variaciones en la estadística anual del TNS, así como detectar anomalías y obtener datos confiables para hacer estudios de física solar. Si bien la variación diurna no es una característica de los rayos cósmicos, se sabe que es uno de los comportamientos que presentan todos los datos registrados por detectores en tierra. Para que los datos del TNS sean estables, su variación debe ser mínima. Como los incrementos o disminuciones en las cuentas deben ser graduales por lo que se utilizará esta variación para ver que los datos del TNS que corresponden a las partículas neutras sean estables, es decir, que no haya observaciones inusuales, y si las hay estudiar su procedencia.

En la figura 4.1 se puede ver la variación diurna registrada el 3 de febrero de 2014 por el TNS en Sierra Negra.

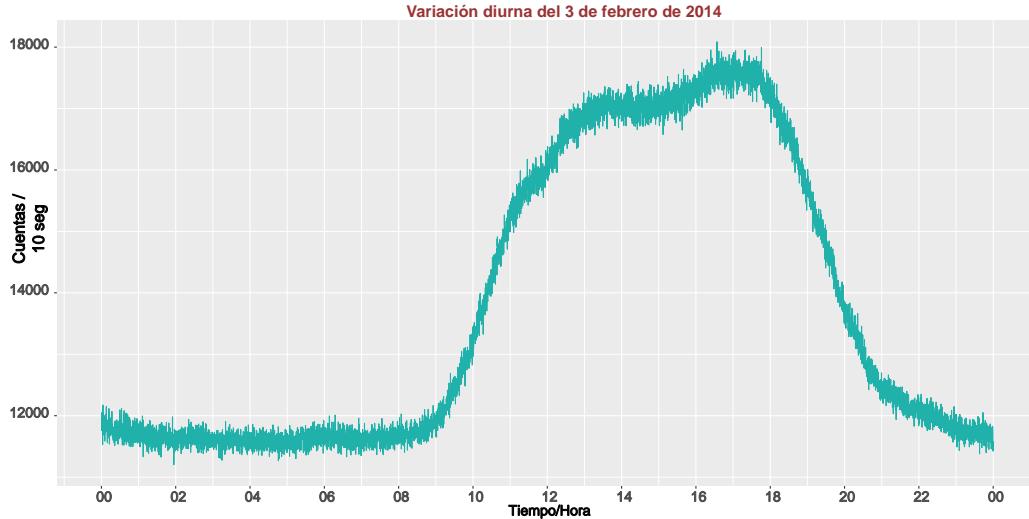


Figura 4.1: Variación diurna registrada por el TNS en Sierra Negra. 3 de febrero de 2014 .

4.2. Estabilidad de la variación diurna

Para comprobar la estabilidad de las observaciones a lo largo del tiempo, las observaciones deben trazarse en función del tiempo. Debido a la gran cantidad de datos del TNS, para detectar patrones se usará la técnica de visualización de datos y se hará sobre la variación diurna, es decir, se verificará que la variación diurna se comporta de manera estable para los 13 años de detección.

A continuación se muestran los gráficos por año para los cuatro canales de energía con anti-coincidencia electrónica.

En la Figura 4.2 se muestra el comportamiento de los datos del canal *S1_A* del año 2004. Para observar el comportamiento de la variación diurna se graficó con razón de conteo 1 minuto. Se puede ver en el gráfico una línea de tendencia (línea roja) y un sombreado que muestra el intervalo de confianza del 95 % para la media, esto último aplica para todos los gráficos de variación temporal.

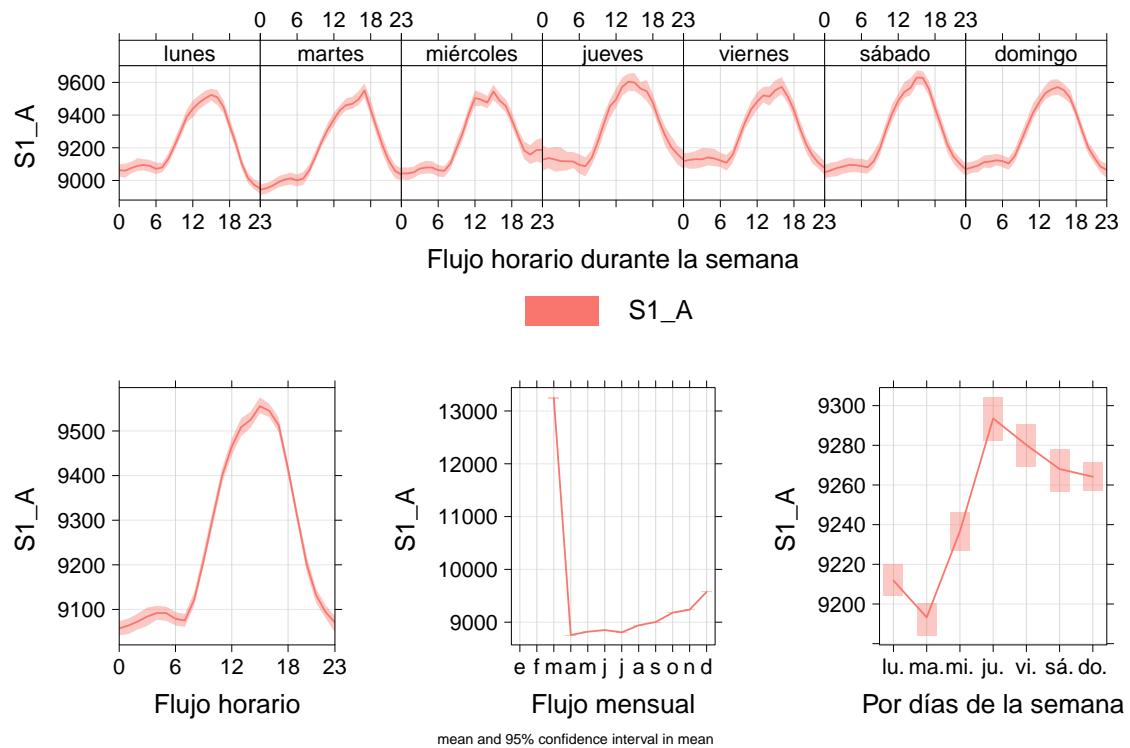


Figura 4.2: Variación temporal de partículas neutras para el canal *S1_A* en 2004

Primero se muestra una evolución semanal distribuidos en horas del día. Se puede observar que hay un máximo cerca de las 18 horas y que la variación diurna se comporta de manera estable durante toda la semana excepto en miércoles, esto último puede deberse a algún error electrónico en alguno de los días miércoles.

En la segunda gráfica se muestra un resumen del comportamiento en horas del día de la variación diurna durante todo el año, en promedio la variación diurna es estadísticamente estable.

El tercer gráfico es un resumen donde se establece la evolución de las cuentas del canal *S1_A* en función de los meses del año. Se puede observar que las cuentas en marzo son muy altas comparadas con el resto de los meses del año. Hay que recordar que para esta fecha aún se estaban haciendo las pruebas para el funcionamiento del TNS.

La última gráfica resume el comportamiento de los datos por días para observar la variación a lo largo de la semana. Se puede notar que en promedio los jueves se tienen las cuentas más altas y de acuerdo con el primer gráfico también se puede observar que en ese día el porcentaje de variación de los datos es mayor pues la banda del intervalo de confianza para la media es más ancha.

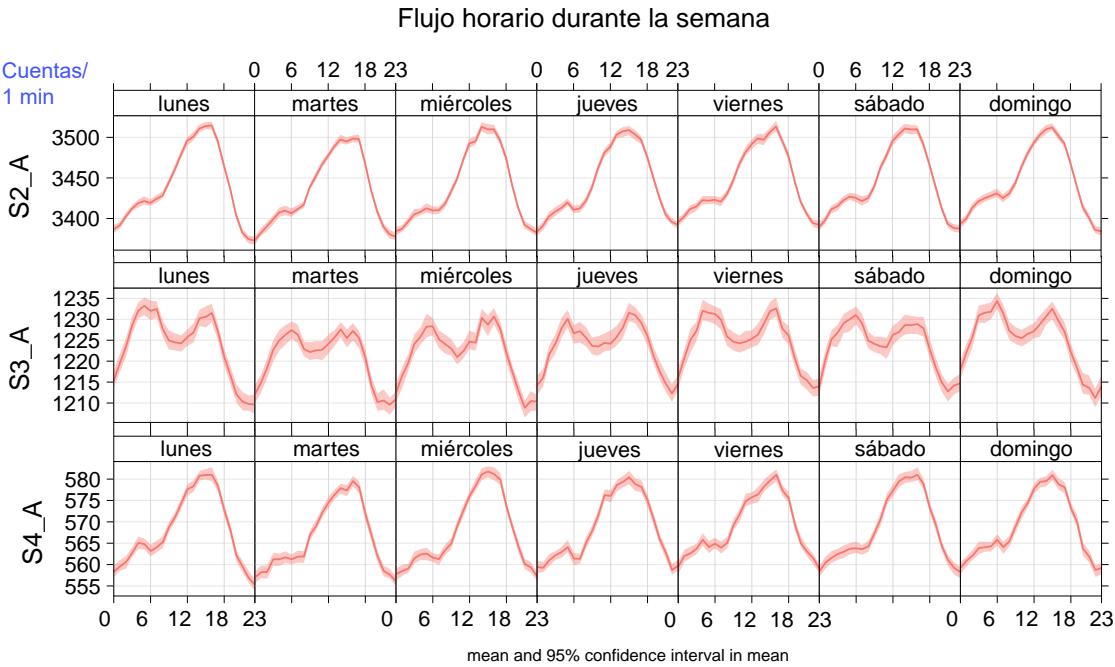


Figura 4.3: Variación de las cuentas en 2004 para los cuatro canales

En la Figura 4.3 se muestra la variación semanal en horas para los canales de energía S2_A, S3_A y S4_A. El segundo y cuarto canal se comportan de forma similar, pero el canal 3 se está comportando de manera diferente en toda la semana, tiene dos picos, es decir, la variación diurna no se muestra estable. Esto puede ser un error electrónico, para ello sería bueno graficar toda la señal del canal 3. En la Tabla 4.1 se puede ver que en la mayoría de los canales de energía, la media de los datos registrados es mayor que la mediana, eso quiere decir que los datos son sesgados a la derecha. Además se puede notar que los datos del canal S1_A tienen un grado de dispersión mayor que los datos de los otros canales.

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	Desv.est	C.V
S1_A	6480	8808	9014	9249	9331	18018	956.68	10.34
S2_A	2383	3361	3434	3440	3514	7120	111.63	3.24
S3_A	830	1192	1223	1223	1254	2692	45.44	3.71
S4_A	397.0	548.0	567.0	567.8	587.0	1311.0	28.87	5.08

Tabla 4.1: Parámetros característicos para los datos del año 2004

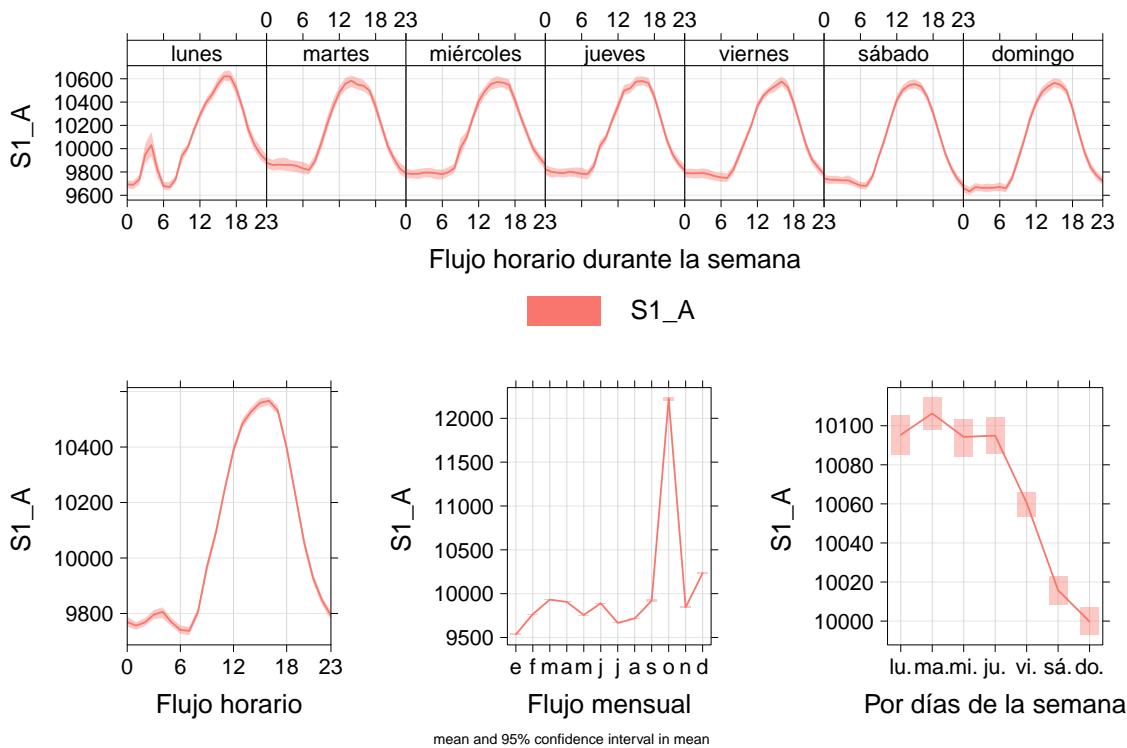


Figura 4.4: Variación de las cuentas para S1_A en 2005

En la Figura 4.4 se puede ver que la variación diurna semanal para S1_A tiene un comportamiento estable excepto el lunes que hay un pico extraño. Solo habría que analizar más los datos para ese día. En la Figura 4.5 se muestran las variaciones semanales para los canales restantes, se puede notar que el pico mencionado para el canal 1 también aparece en éstos; lo que hace pensar que no puede ser un error electrónico pero hay que ser cuidadosos y hacer más análisis. Además se puede observar que el canal 3 sigue estando muy inestable. De acuerdo a las medidas de la Tabla 4.2 se puede decir que los datos están sesgados a la derecha y su variabilidad fue aumentando para cada uno de los canales de energía.

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	Desv.est	C.V (%)
S1_A	6603	9451	9754	10067	10300	31803	1180.71	11.72
S2_A	976	3579	3731	3762	3863	13151	691.62	18.38
S3_A	286	1254	1317	1335	1366	5204	348.16	26.07
S4_A	129.0	582.0	622.0	635.5	656.0	2567.0	187.44	29.49

Tabla 4.2: Parámetros característicos para los datos del año 2005

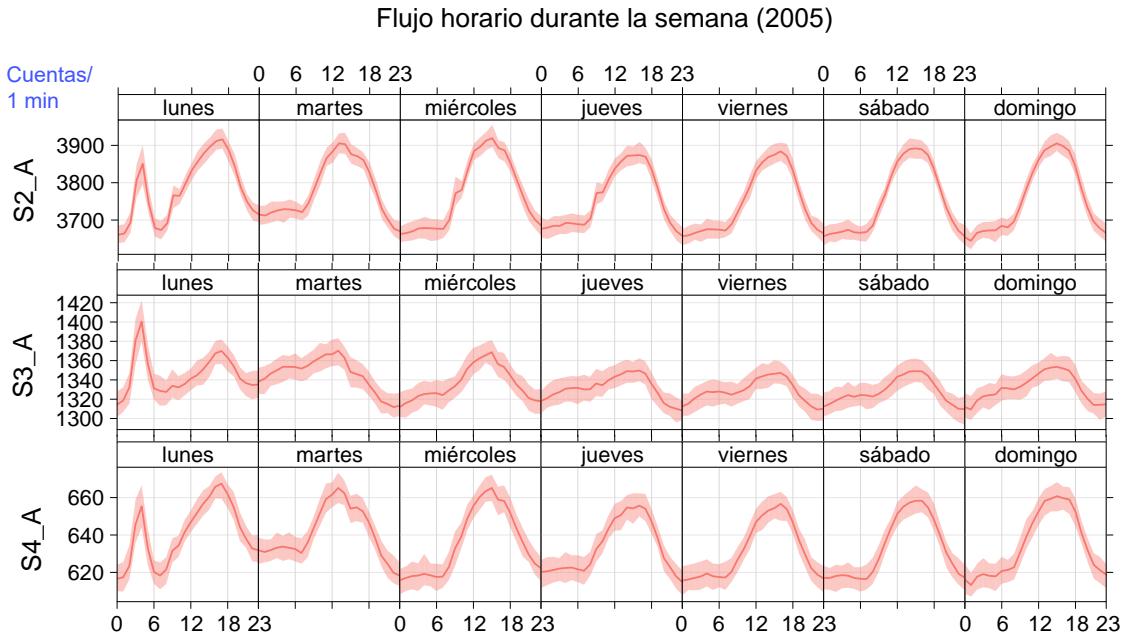


Figura 4.5: Variación de las cuentas en 2005 para los cuatro canales

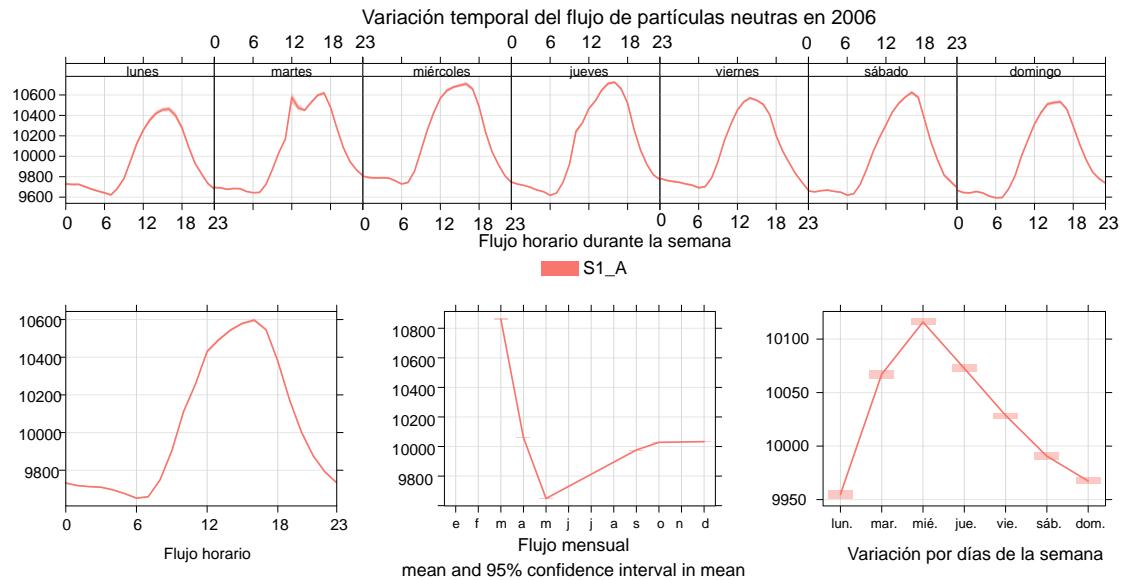


Figura 4.6: Variación de las cuentas para $S1_A$ en 2006

En las figuras 4.6 y 4.7 se muestran las gráficas de variación para el año 2006, se observa que la variación diurna se comporta de manera estable en los cuatro canales de energía. Aunque hay un pico en el día martes, se sabe que en este comportamiento no influye el cambio de razón de conteo. En la Tabla 4.3 se puede ver que la dispersión de los datos es mínima para

todos los canales de energía, entonces el pico se produjo de manera gradual y hay que hacer más análisis en ese intervalo.

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	Desv.est	C.V. (%)
S1_A	8154	9630	9905	10033	10313	26476	624.45	6.22
S2_A	2974	3349	3455	3465	3562	11289	183.22	5.28
S3_A	956	1108	1137	1139	1168	4321	49.88	4.37
S4_A	404.0	507.0	528.0	530.2	551.0	2131.0	35.29	6.65

Tabla 4.3: Parámetros característicos para los datos del año 2006

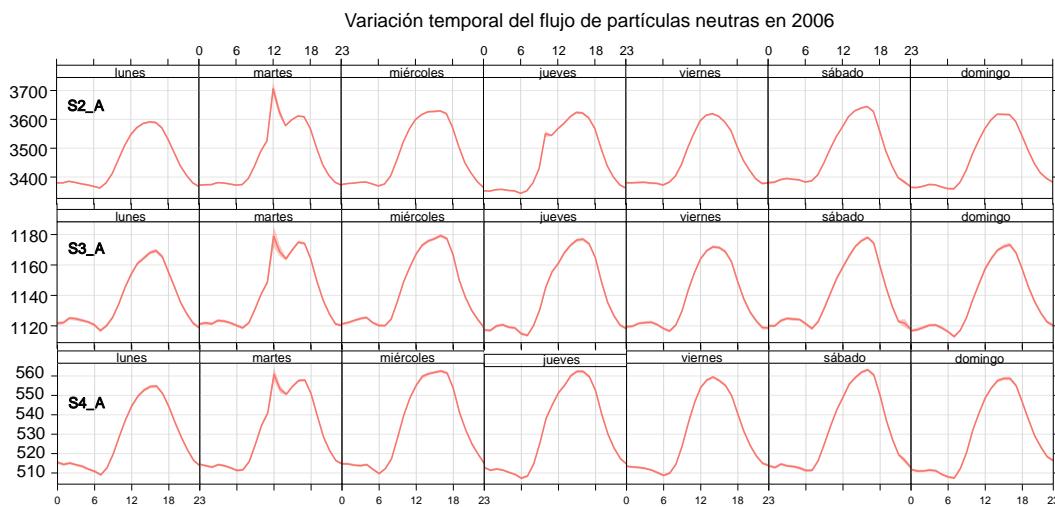


Figura 4.7: Variación diurna en 2006 para los cuatro canales con anti

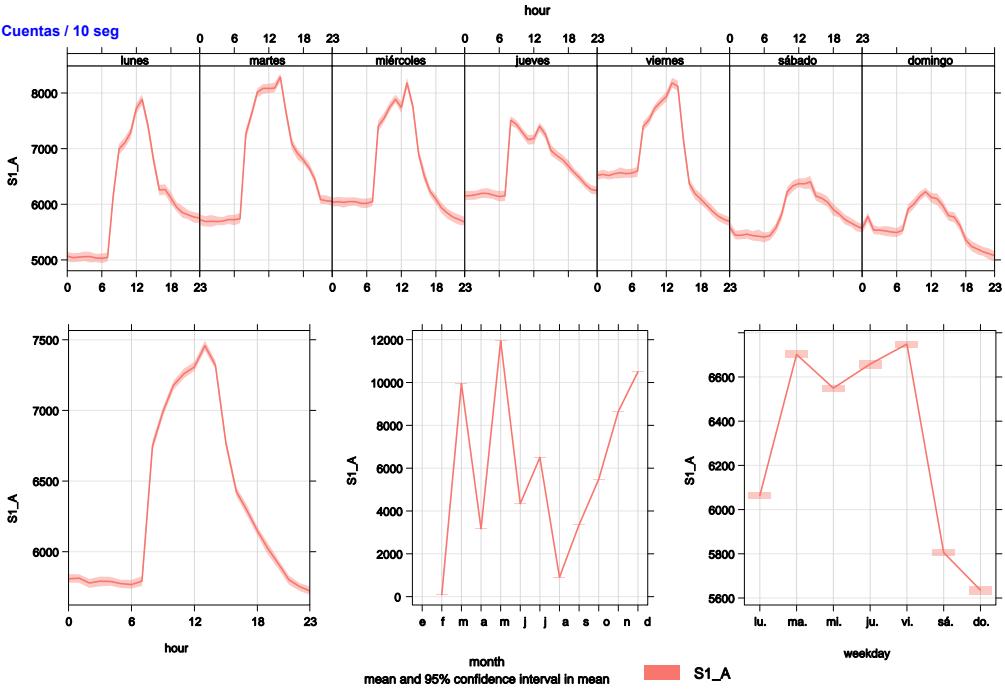


Figura 4.8: Variación de las cuentas para $S1_A$ en 2008

En el año 2008 el TNS volvió a conectarse y en marzo las cuentas fueron sobre 3 segundos, así que para ver la variación de esta señal se hizo un incremento del tiempo de 10 segundos. En las figuras 4.8 y 4.9 se puede ver la tendencia de la señal en este año. Se puede concluir que este año es el más inestable, definitivamente la variación diurna no se está comportando de la manera esperada en ninguno de los canales. La Tabla 4.4 muestra medidas de centralización y dispersión para $S1_A, \dots, S4_A$ de datos registrados cada 10 segundos. Se puede notar que los tres primeros canales muestran una variabilidad por debajo del 18 %, mientras que los datos del último canal presentan un alto grado de dispersión. Además, el 25 % de estos datos se encuentran entre 92 y 1148, al hacer más análisis se encontró que menos del 1 % se encuentra por arriba de 1000 y el resto por debajo de 600. Por lo que al utilizar los datos para diferentes análisis físicos, se tienen que analizar para discernir de las posibles influencias de los errores electrónicos, en la toma de datos o de software.

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	Desv.est	C.V (%)
S1_A	9216	10304	11292	11357	12344	25724	1309.05	11.52
S2_A	3072	3128	3184	3634	4152	10256	636.01	17.49
S3_A	1024	1056	1084	1087	1116	3196	65.18	5.99
S4_A	4.00	28.00	56.00	61.13	92.00	1148.00	54.66	89.42

Tabla 4.4: Parámetros característicos para los datos del año 2008

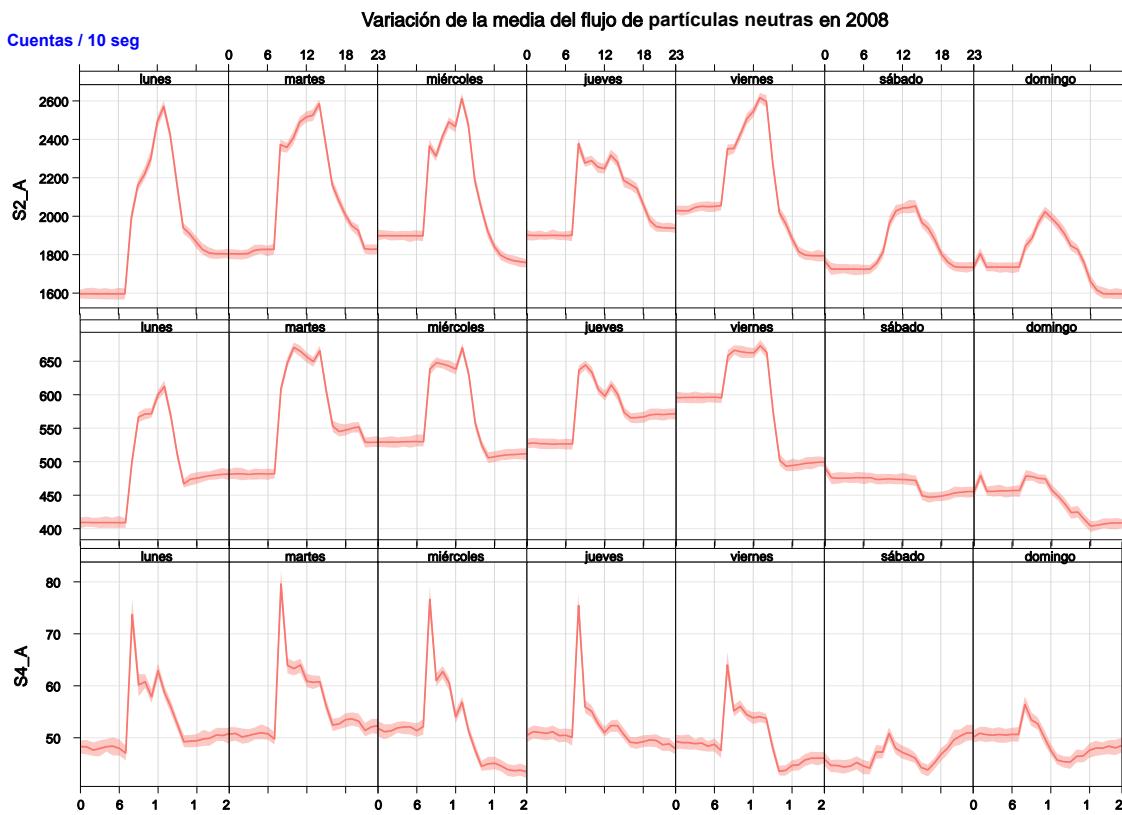


Figura 4.9: Variación diurna en 2008 para los cuatro canales con anti

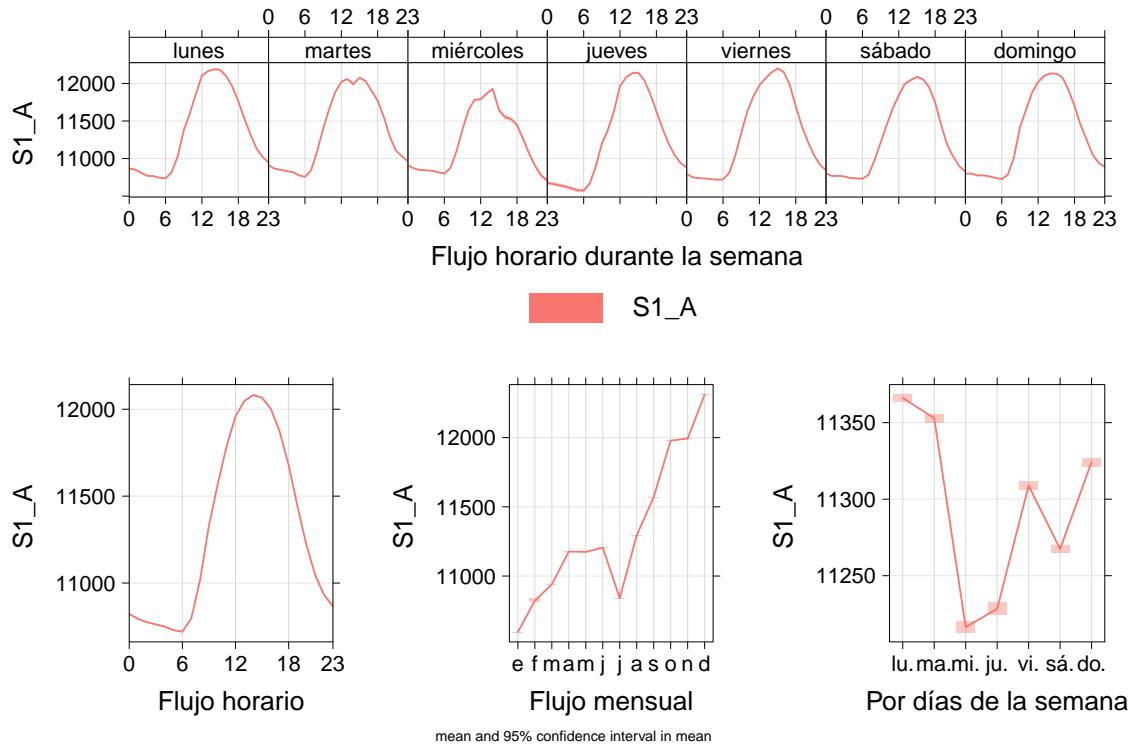


Figura 4.10: Variación de las cuentas para $S1_A$ en 2009

Para el año 2009 se puede ver en las figuras 4.10 y 4.11 que la variación diurna se comporta de manera estable en los canales 1, 2 y 3, lo cual no se puede decir lo mismo del cuarto canal. Hay que notar también que entre miércoles y jueves se produce una caída en las cuentas. Al hacer más análisis se encontró que un poco más del 50 % de los datos del canal S4_A se había generado por errores electrónicos y eso causaba que en la gráfica se mostrara una banda de intervalo de confianza muy grande. En la tabla 4.5 se puede ver que los datos del canal S1_A y S4_A presentan casi el mismo grado de variabilidad.

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	Desv.est	C.V
S1_A	7037	10348	11324	11325	11832	22648	849.86	7.5
S2_A	2048	3156	4120	3830	4215	8216	567.7	14.82
S3_A	809	1072	1124	1194	1343	3124	148.92	12.47
S4_A	239.0	611.0	638.0	644.8	672.0	1379.0	48.32	7.49

Tabla 4.5: Parámetros característicos para los datos del año 2009

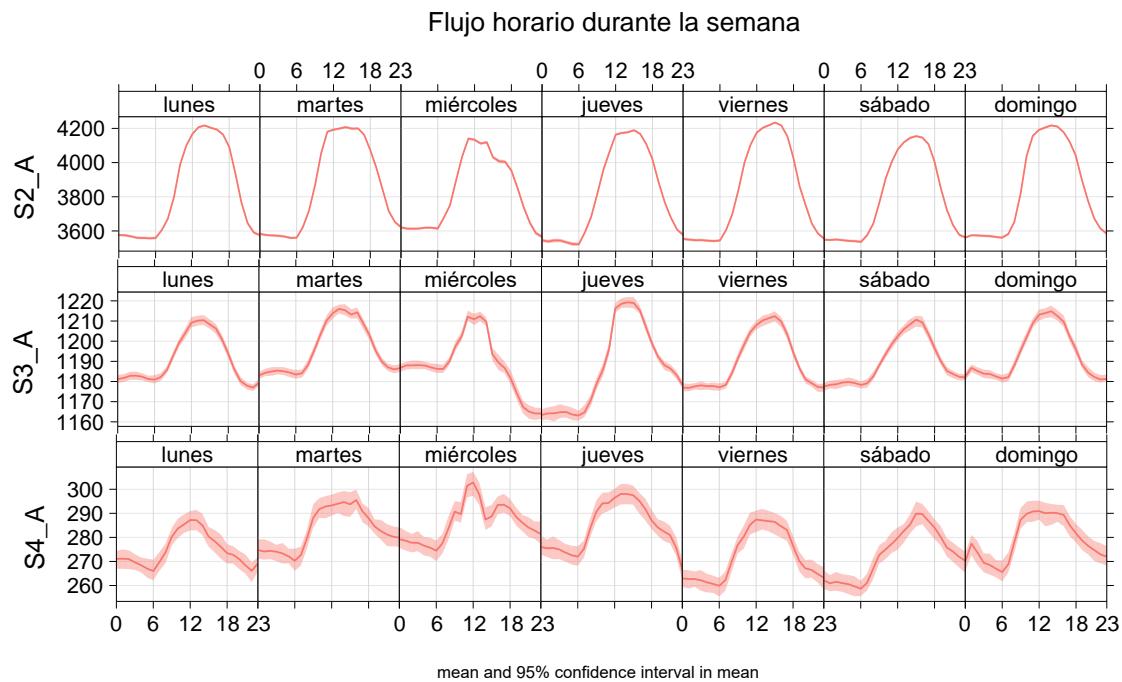


Figura 4.11: Variación diurna en 2009 para los cuatro canales con anti

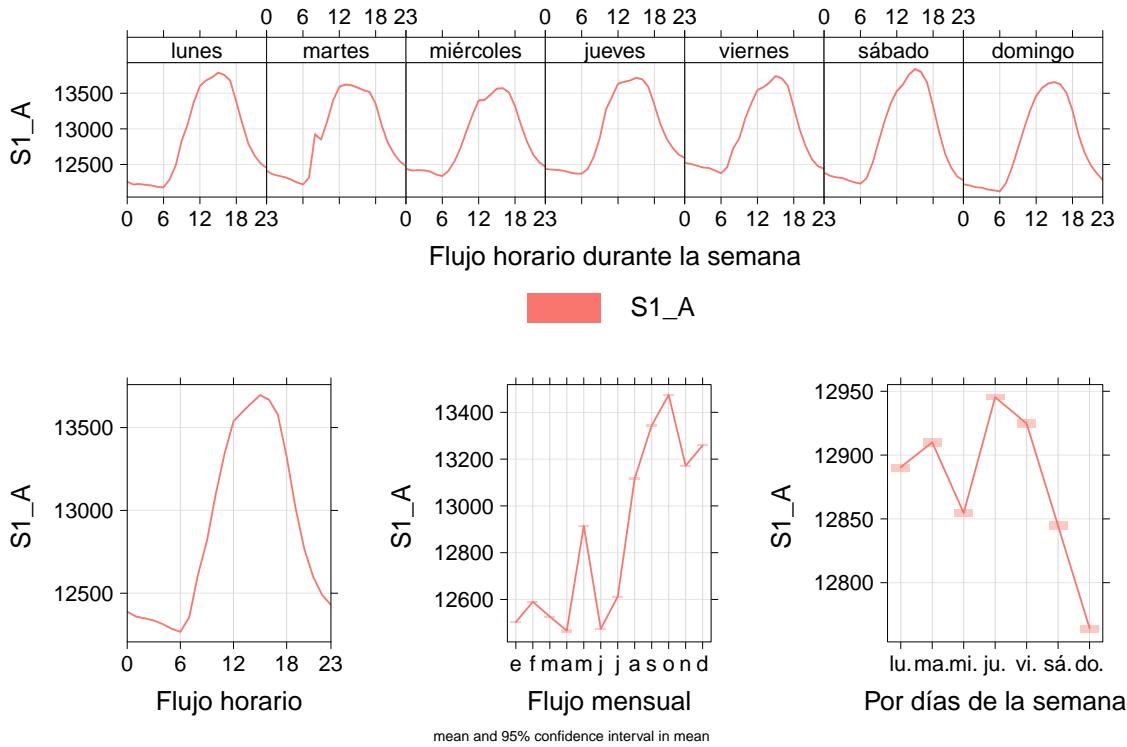


Figura 4.12: Variación de las cuentas para $S1_A$ en 2010

En el año 2010 en la figura 4.12 y 4.13 se puede observar que el canal 1 se está comportando de manera estable, pero en los demás canales hay un pico en los días martes y miércoles. Los datos para todas las variables ($S1_A, \dots, S4_A$) siguen estando ligeramente sesgados a la derecha y el porcentaje de dispersión está por debajo del 8.4 %.

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	Desv.est	C.V
$S1_A$	8834	12266	12741	12878	13354	26629	855.99	6.64
$S2_A$	2905	4527	4689	4734	4895	10749	300.46	6.34
$S3_A$	950	1426	1475	1485	1532	3705	85.61	5.76
$S4_A$	433.0	657.0	690.0	697.6	730.0	1874.0	58.41	8.37

Tabla 4.6: Parámetros característicos para los datos del año 2010

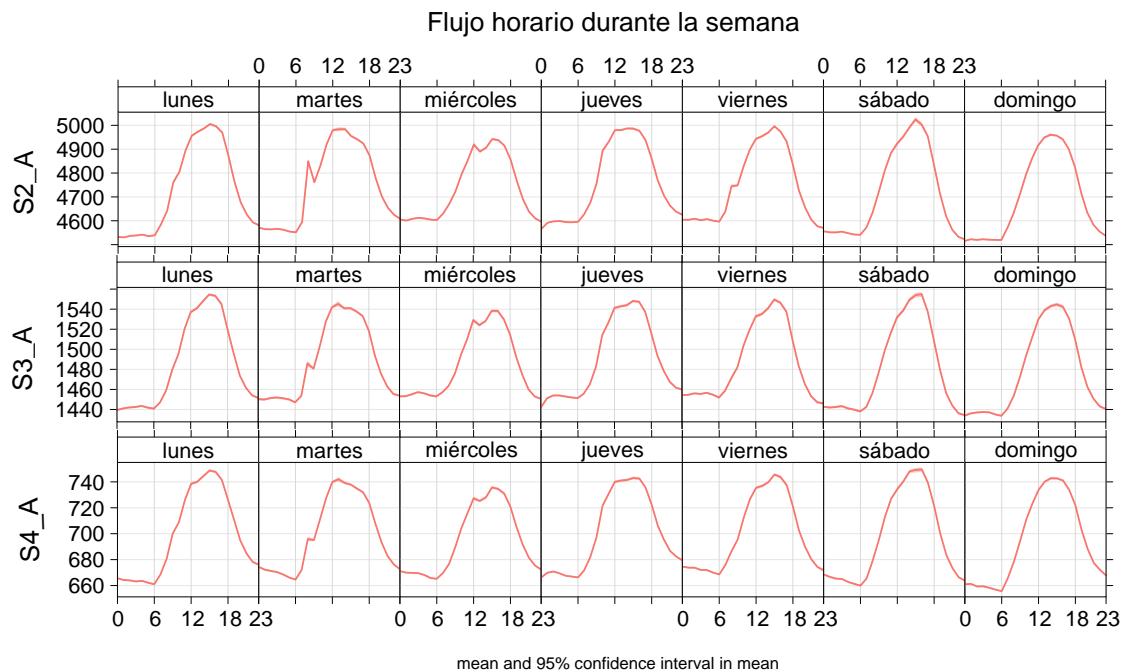


Figura 4.13: Variación diurna en 2010 para los cuatro canales con anti

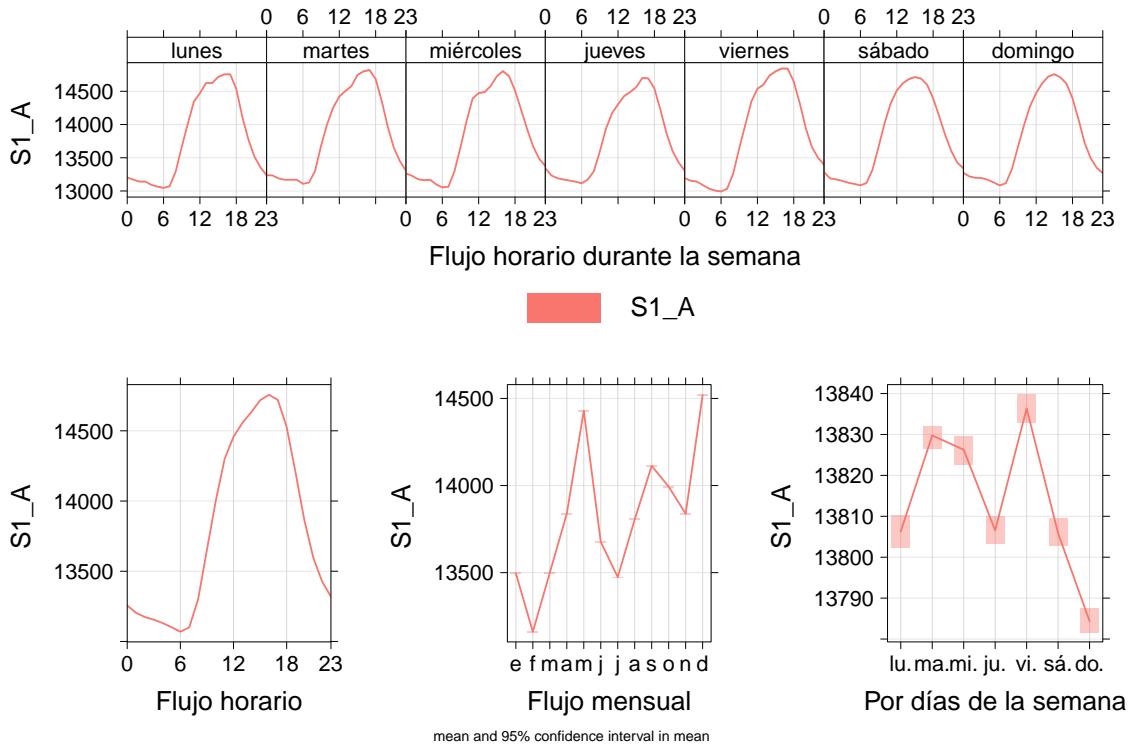


Figura 4.14: Variación de las cuentas para $S1_A$ en 2011

En las figuras 4.14 y 4.15 se muestran las variaciones semanales para el año 2011. Se observa que las variaciones son estables, ligeros picos y de acuerdo a la banda de confianza para la media y el coeficiente de variación (ver Tabla 4.7) los datos tienen un grado de dispersión bajo, entre 6.1% y 8.8%. En la variación mensual se nota que las cuentas van creciendo, donde mayo y diciembre registran las variaciones más altas.

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	Desv.est	C.V
S1_A	9023	13091	13667	13813	14401	26350	972.88	7.04
S2_A	2945	4960	5153	5202	5398	11687	349.04	6.7
S3_A	1011	1568	1629	1639	1699	4494	100.81	6.15
S4_A	456	733	774	782	823	2152	68.3	8.73

Tabla 4.7: Parámetros característicos para los datos del año 2011

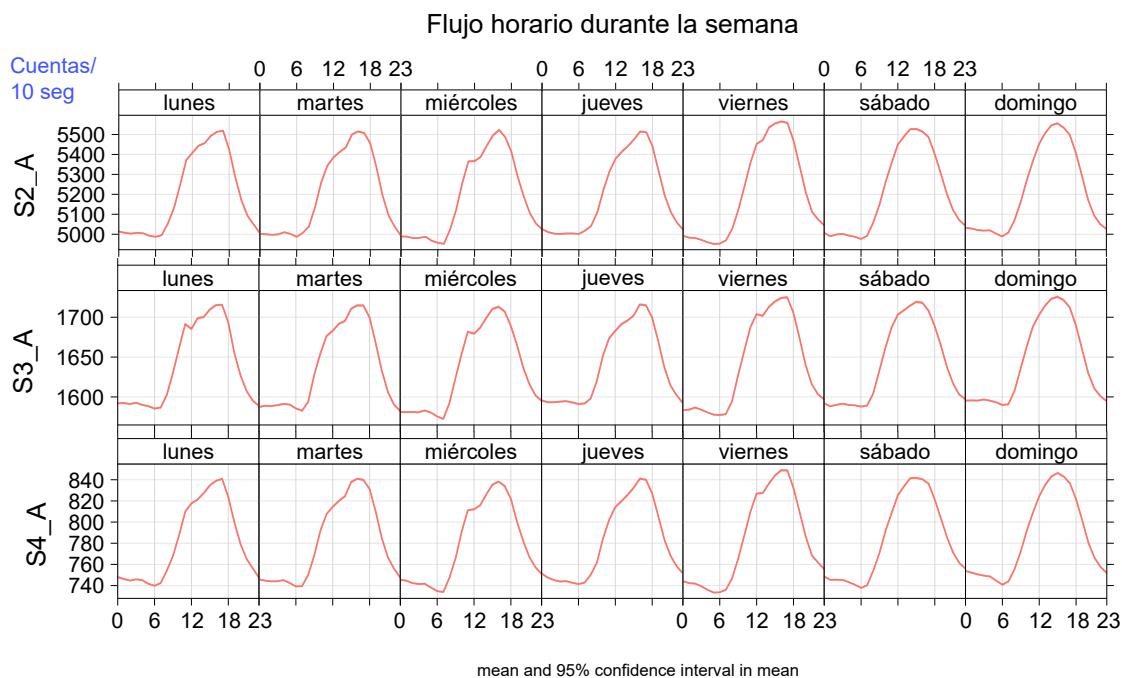


Figura 4.15: Variación de las cuentas para los cuatro canales en 2011

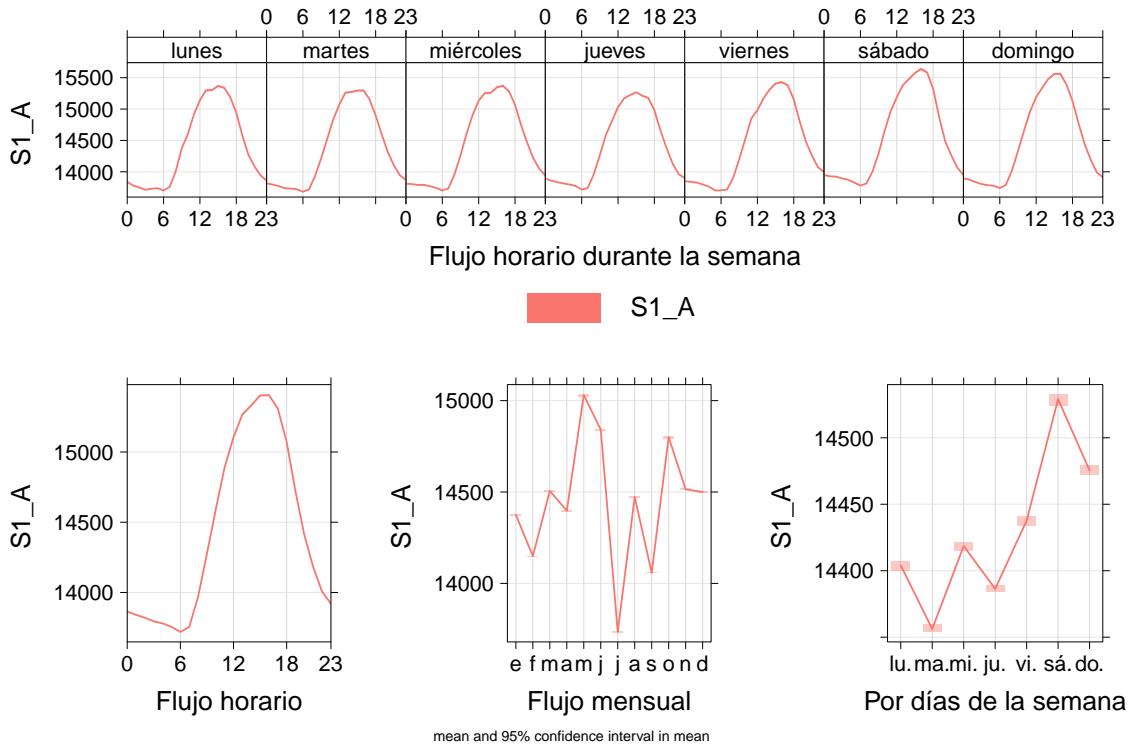


Figura 4.16: Variación de las cuentas para $S1_A$ en 2012

Se observa en las figuras 4.16 y 4.17 que la variación diurna del año 2012 para los tres primeros canales de energía con anticoincidencia electrónica es estable, con una variación en los datos muy pequeña. Aunque este comportamiento no es similar en el cuarto canal, además de tener picos, la variación está trasladada, lo que sugiere que existen errores en la toma de datos y/o software. Los datos para este año están sesgados a la derecha y el canal $S4_A$ presentando un porcentaje de variación mayor comparado con los datos de otros canales (ver Tabla 4.8).

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	Desv.est	C.V
$S1_A$	8214	13790	14186	14430	14859	46461	1020.12	7.06
$S2_A$	3170	5428	5604	5650	5811	12871	332.17	5.87
$S3_A$	972	1706	1760	1773	1825	5001	98.87	5.57
$S4_A$	216.0	742.0	785.0	769.7	817.0	2413.0	72.83	9.46

Tabla 4.8: Parámetros característicos para los datos del año 2012

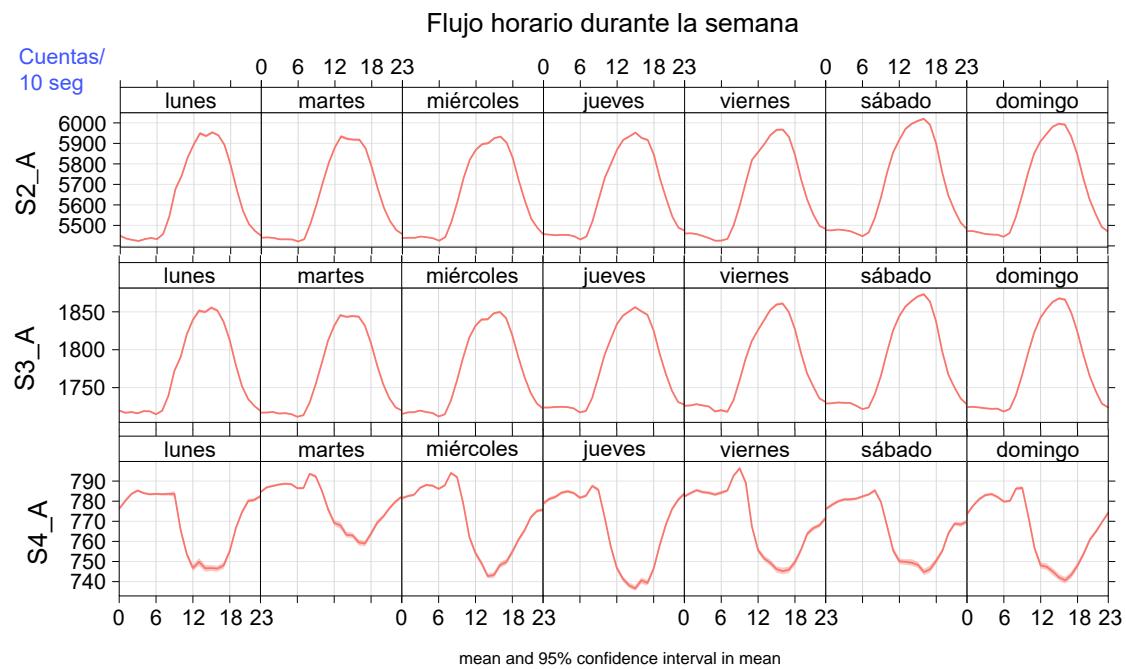


Figura 4.17: Variación diurna en 2012 para los cuatro canales con anti

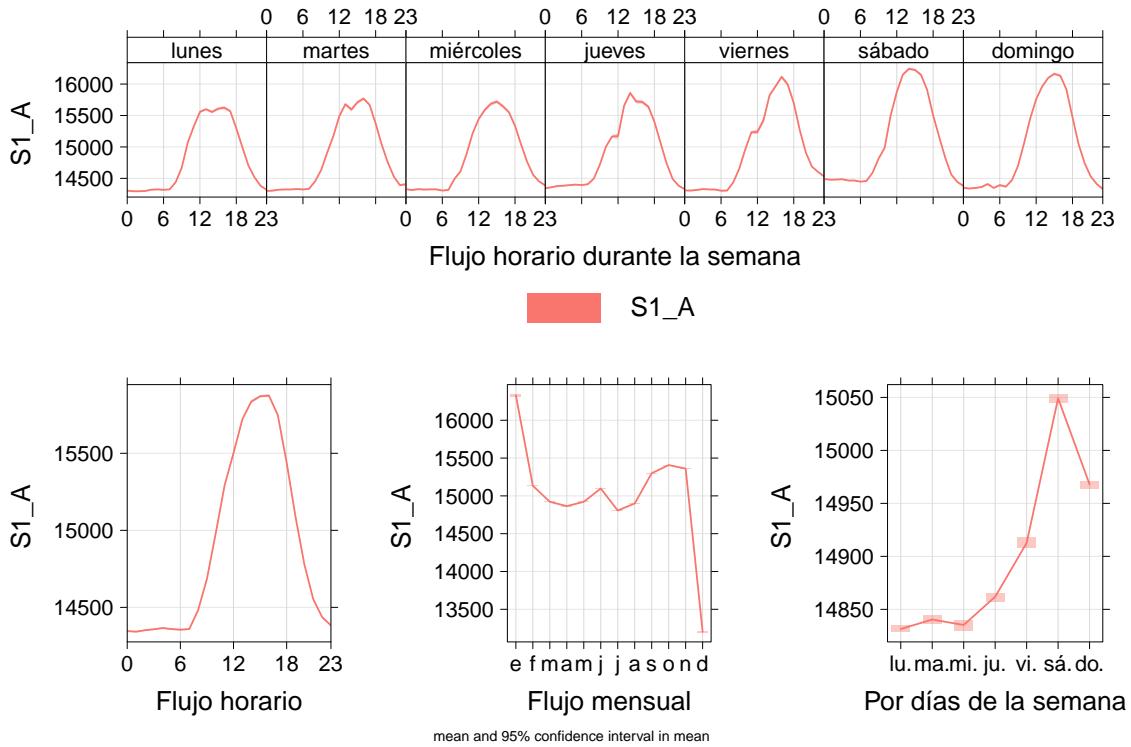


Figura 4.18: Variación de las cuentas para $S1_A$ en 2013

En la figura 4.18 y 4.19, gráficas que corresponden a datos del año 2013, se observa que los tres primeros canales tienen un comportamiento similar con picos que se notan más en el canal 2 y 3, estos picos se generaron por errores electrónicos que fueron muy recurrentes en este año. En el canal 4 se observa una variación diurna más estable salvo que está trasladada. Además, se puede observar en la Tabla 4.9 que los datos de este último canal están sesgados a la izquierda, no a la derecha como los otros, y el grado de dispersión mayor que en las otras variables. Por lo tanto, aún cuando la variación diurna esté trasladada, los datos del $S1_A$ son los mejores ya que se comportan de la manera esperada.

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	Desv.est	C.V
$S1_A$	7006	14418	14814	14908	15352	28663	1277.88	8.57
$S2_A$	1920	5830	5991	5965	6156	11695	336.16	5.63
$S3_A$	702	1799	1849	1840	1901	4233	106.02	5.76
$S4_A$	236.0	603.0	702.0	693.6	785.0	2122.0	104.95	15.13

Tabla 4.9: Parámetros característicos para los datos de 2013

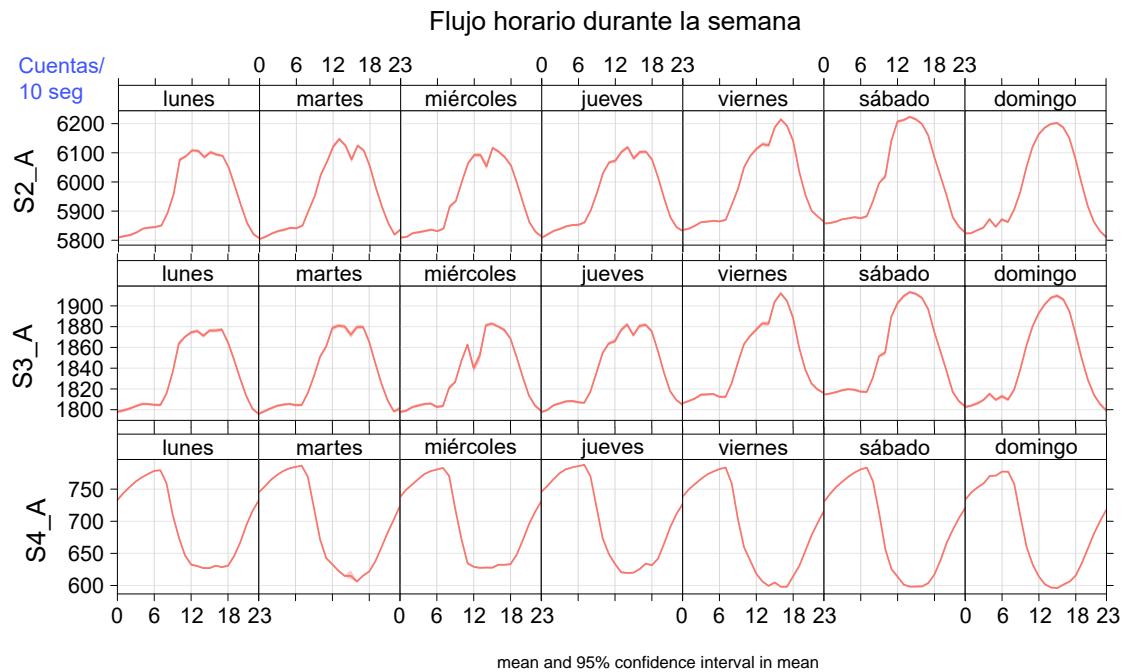


Figura 4.19: Variación diurna en 2013 para los cuatro canales con anti

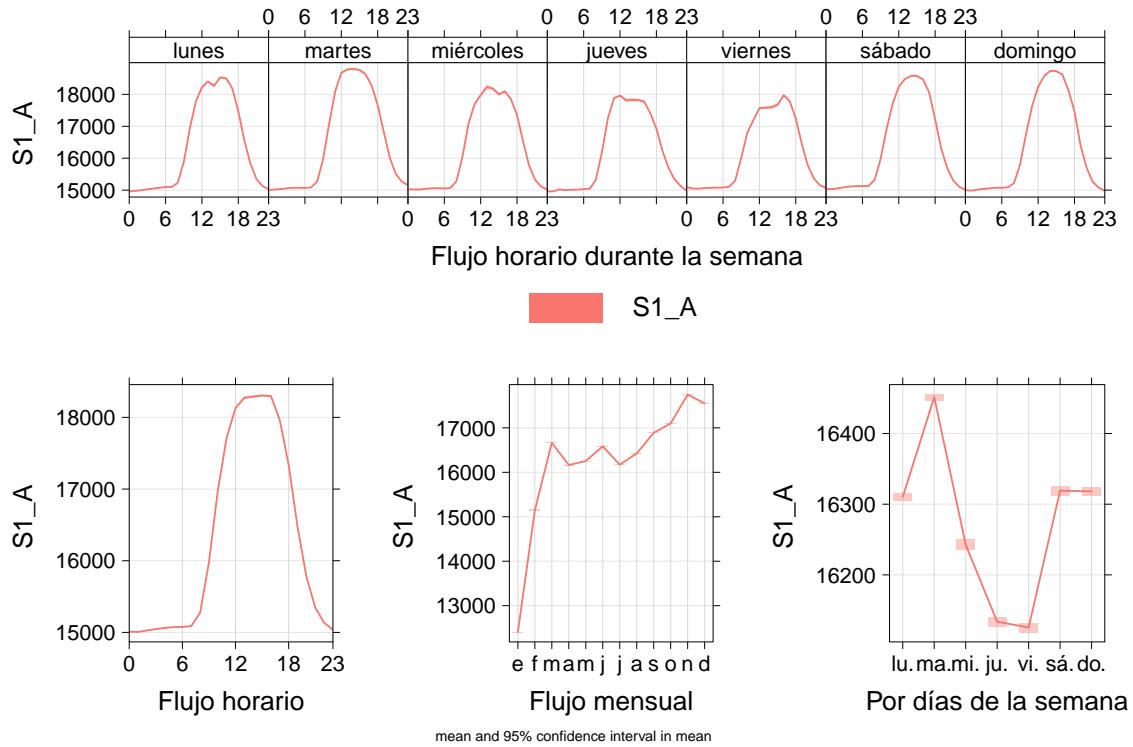


Figura 4.20: Variación de las cuentas para $S1_A$ en 2014

En la figura 4.20 y 4.21 se muestran las variaciones semanales por horas del año 2014. En el canal 1 se observan ligeros picos, los cuales se acentúan en los canales 2 y 3, mientras que el cuarto canal sigue mostrando una variación diurna trasladada. El intervalo de confianza apenas se nota, eso indica que la variación es pequeña. En la Tabla 4.10 se puede notar que el canal $S1_A$ y $S4_A$ son los que presentan mayor grado de variabilidad comparado con los datos de otros canales.

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	Desv.est	C.V
$S1_A$	7154	15348	15817	16292	17149	29284	2322.95	14.25
$S2_A$	1003	6259	6453	6327	6605	11840	483.92	7.64
$S3_A$	712	1901	1961	1928	2013	4371	145.29	7.53
$S4_A$	206.0	525.0	578.0	596.1	664.0	2761.0	105.94	17.77

Tabla 4.10: Parámetros característicos para los datos del año 2014

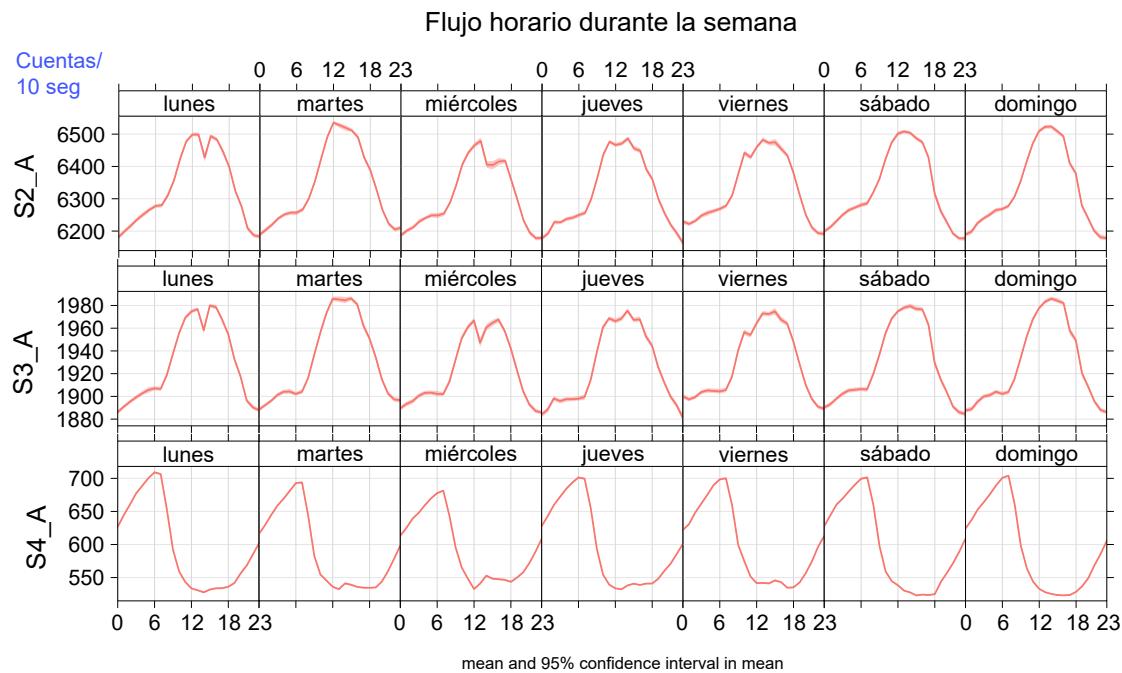


Figura 4.21: Variación diurna en 2014 para los cuatro canales con anti

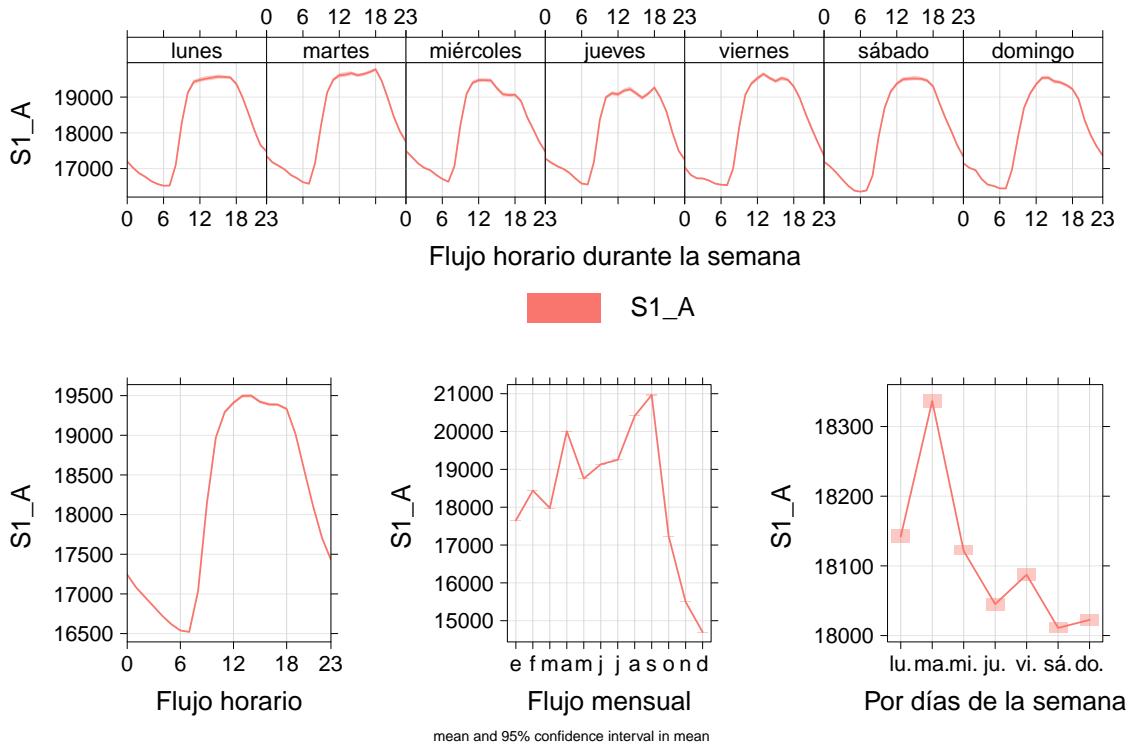


Figura 4.22: Variación de las cuentas para $S1_A$ en 2015

La Figura 4.22 muestra una variación diurna ligeramente trasladada para el año 2015, mientras que en la Figura 4.23 se observa que el canal 2 y 3 ahora muestran la variación diurna estable con un porcentaje de variación de los datos menor al 6.8 % (ver Tabla 4.11). El cuarto canal muestra picos y datos trasladados que se deben a errores electrónicos.

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	Desv.est	C.V
S1_A	7839	16041	16842	18110	20963	27974	2861.88	15.8
S2_A	3525	6573	6761	6916	7238	13007	464.18	6.71
S3_A	1104	2003	2064	2077	2138	4938	103.58	4.98
S4_A	234.0	431.0	455.0	466.7	485.0	1961.0	61.09	13.09

Tabla 4.11: Parámetros característicos para los datos del año 2015

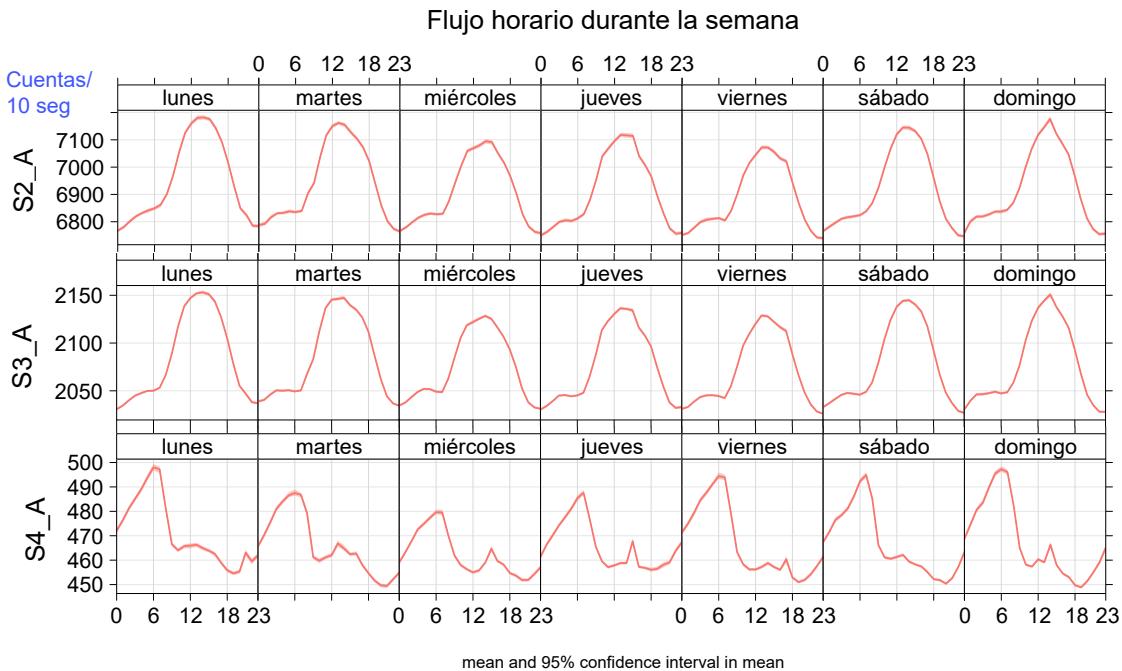


Figura 4.23: Variación diurna en 2015 para los cuatro canales con anti

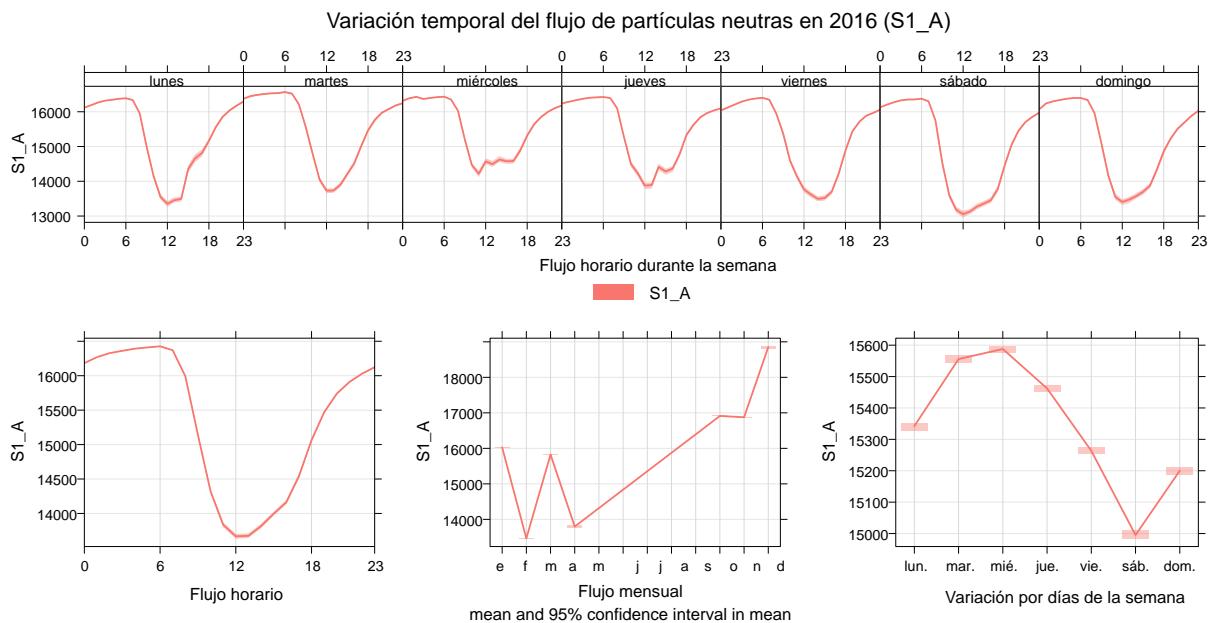


Figura 4.24: Variación de las cuentas para $S1_A$ en 2016

Ahora, para el año 2016, la figura 4.24 muestra una variación diurna estable pero trasladada a 12 horas aproximadamente, esto se debe a que en el año 2015 la variación ya se estaba

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	Desv.est	C.V
S1_A	7937	15394	16077	15340	16469	36883	2266.52	14.77
S2_A	6847	7494	7649	7689	7849	14429	278.95	3.62
S3_A	1911	2152	2205	2223	2277	4536	111.46	5.01
S4_A	365.0	466.0	488.0	491.8	512.0	3151.0	55.11	11.2

Tabla 4.12: Parámetros característicos para los datos del año 2016

trasladando debido a errores electrónicos, lo cual provocó que en este año se trasladara aún más. Por otro lado, se puede notar en la gráfica del flujo mensual que la variación ha aumentado en los últimos meses. De acuerdo a la Tabla 4.12, los datos del S1_A están sesgados a la izquierda y presentan un mayor porcentaje de dispersión comparado con los demás canales de energía.

Al analizar los resultados en las gráficas, se puede observar que a lo largo de los 13 años, el canal S1_A tiene un comportamiento estable mientras que los otros canales tienen varios errores electrónicos. En algunos años se observó una variación diurna estable pero con datos trasladados en el tiempo. Por otro lado, se observó que en el 2008 los errores de toma de datos y electrónicos fueron mayores, lo que se puede atribuir a que el equipo estuvo apagado durante 2007, después de ser afectado por fuertes tormentas eléctricas en la cima de Sierra Negra. Con base en las gráficas, se observa que a partir de 2009 la señal se estabilizó.

Se ha mencionado que la variación diurna registra un máximo alrededor de las 18 horas en tiempo local. Veamos qué sucede con los máximos diarios de los datos del TNS en Sierra Negra durante los 13 años. Se observa en la figura 4.25 que los máximos se concentran entre las 12 y 18 horas en tiempo local. Obsérvese también la figura 4.26 que muestra los mismos datos que la figura anterior pero ahora diferenciando por año, se nota que los máximos están aumentando año con año.

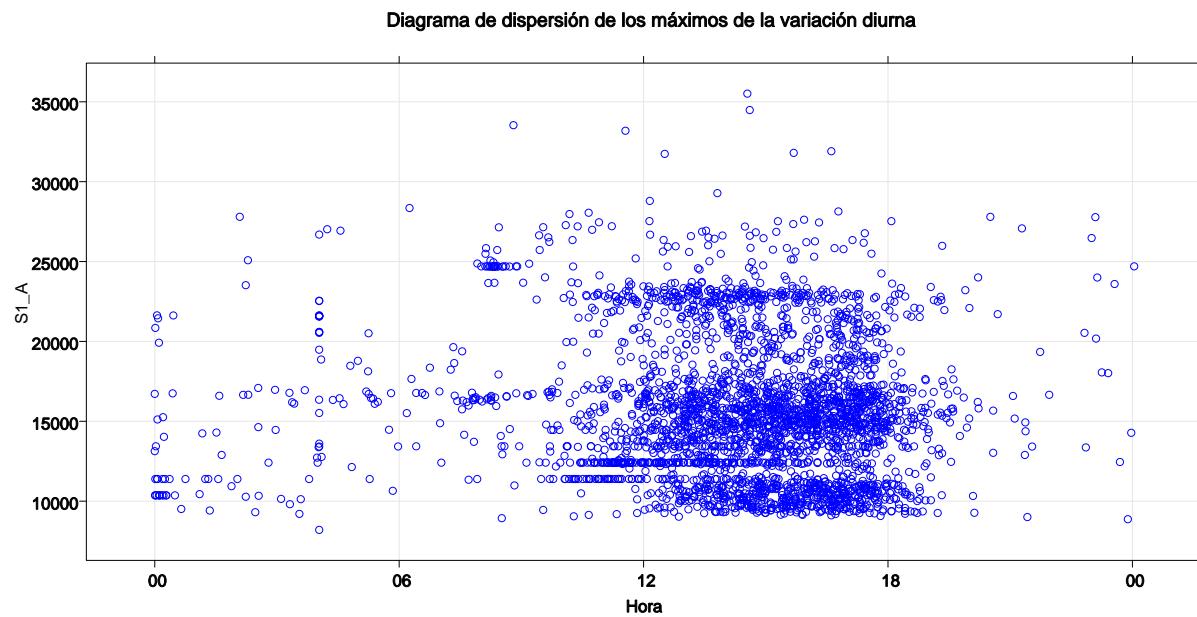


Figura 4.25: Comportamiento de los máximos de la variación diurna

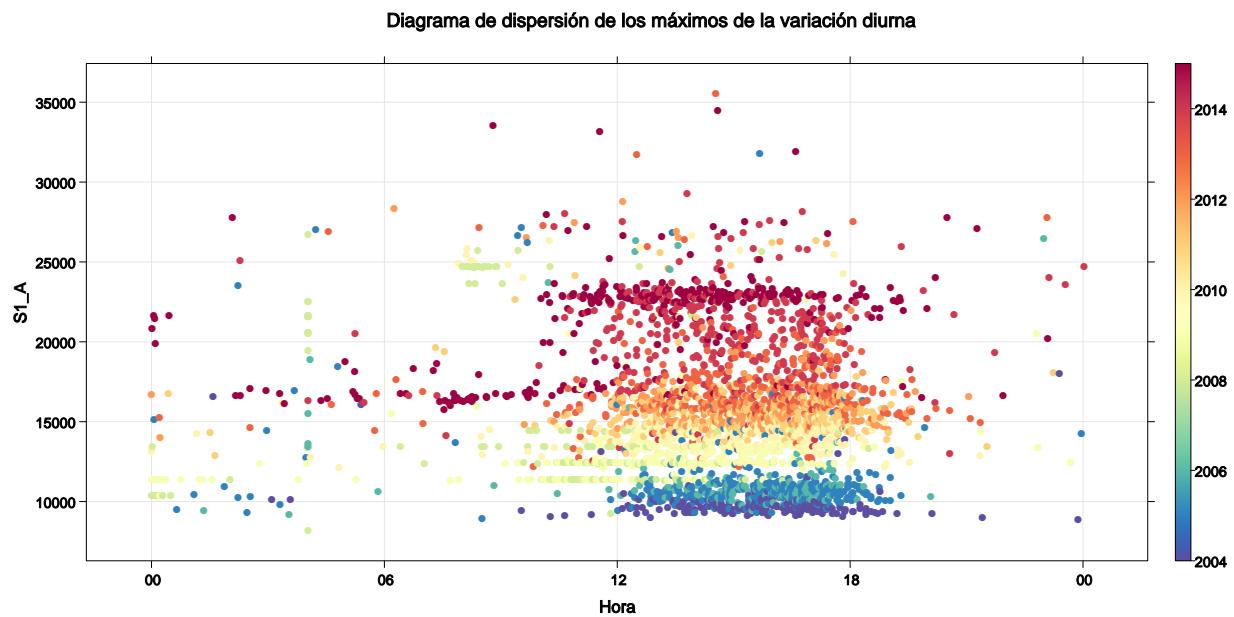


Figura 4.26: Comportamiento por año de los máximos de la variación diurna

4.3. Señal total del TNS

Con base en el análisis de la serie de 13 años de datos del TNS, se obtuvieron el porcentaje de datos confiables por año (figura 4.27) y la señal total del detector (figura 4.28). En la figura 4.27 se muestran los porcentajes de datos estables del TNS en Sierra Negra para cada año; además, con base en los análisis anteriores, observamos que en los últimos años la variación es mínima, ya que el intervalo de confianza se mostró muy bajo. Se observa que desde 2009 el TNS tiene porcentajes altos de detección y se han reducido los errores en la toma de datos, de hardware y software. La figura 4.28 muestra la señal total del TNS, se observa que el flujo de partículas detectadas se ha incrementado. La señal del TNS tiene una tendencia creciente, que es consistente con que el Sol se encuentra rumbo al mínimo de su actividad y las emisiones son menos frecuentes. Este resultado está en coincidencia con el análisis anterior, donde se observa que los máximos de la variación diurna han aumentando cada año, lo que asegura que el número de cuentas registradas también se ha incrementado.

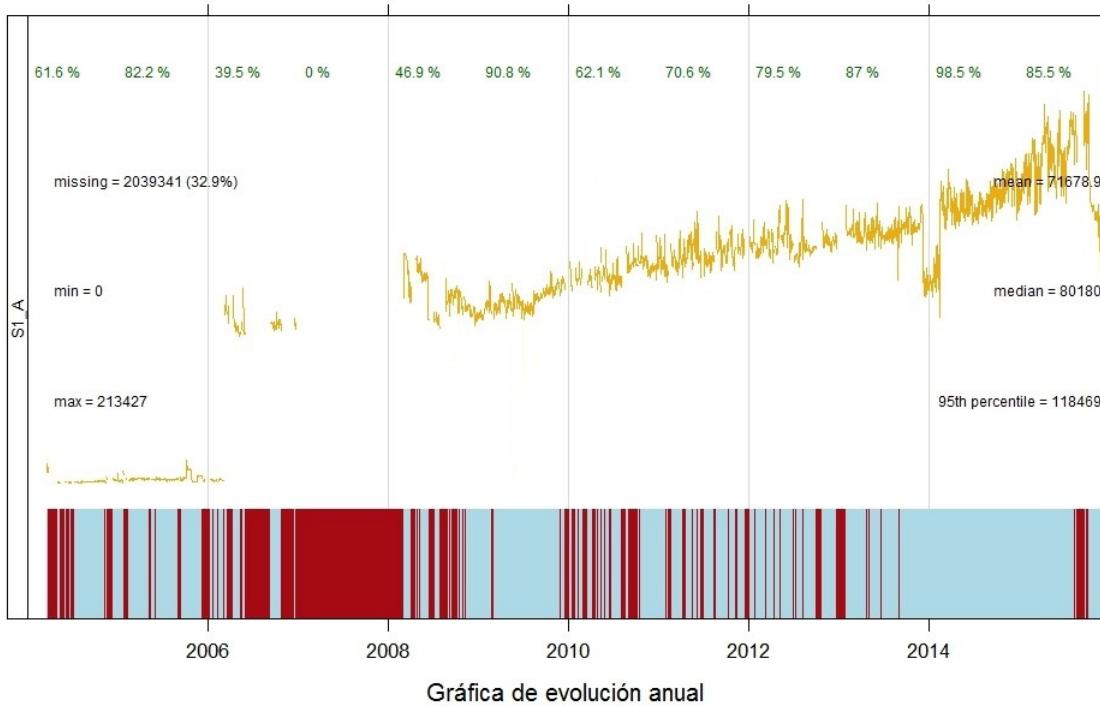


Figura 4.27: Datos para el canal *S1_A* con razón de conteo 1 min. La gráfica muestra un resumen del comportamiento del flujo de partículas neutras del TNS. En la parte de arriba se muestra el porcentaje de datos confiables para cada año. Además da el porcentaje global de datos faltantes (32 %) una cantidad aceptable tomando en cuenta que en el año 2007 hay 0 % de datos confiables pues no hubo registro en ese año. También se muestra una media de 71678.9 (razón de conteo de 1 minuto), mediana igual a 80180 y maximo igual a 213427, nótese que el percentil 95 es de 118469, con estos datos se puede ver que la distribución es asimétrica izquierda pues la mediana es mayor que la media y el 95 % de los datos se concentra de ese lado. En la parte de abajo, con sombreado rojo, se muestra la distribución de datos faltantes a lo largo de 13 años.

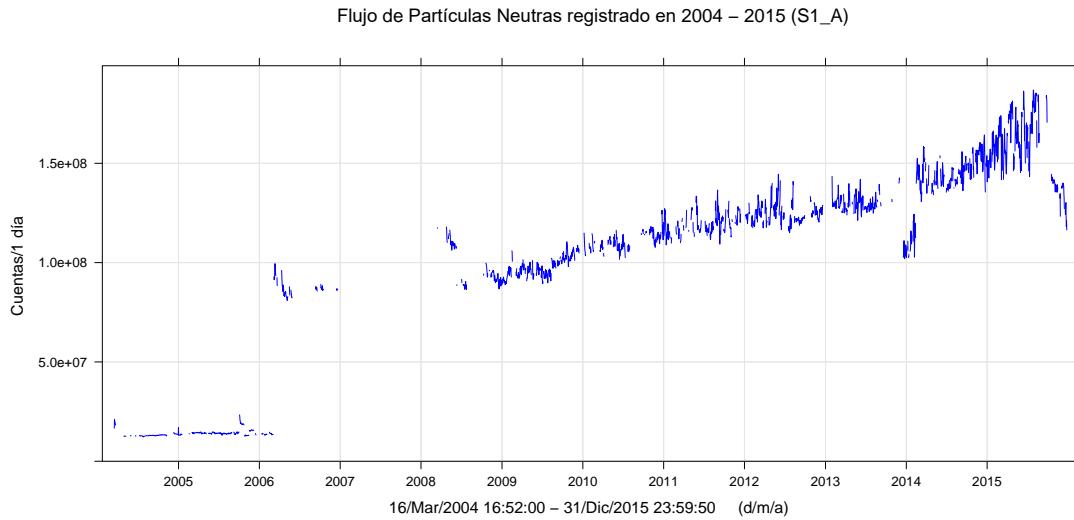


Figura 4.28: Señal completa para $S1_A$ con una razón de conteo de 24 horas. En esta gráfica se puede notar el incremento del flujo de partículas neutras registradas antes del 7 de marzo de 2006 y después de esta fecha, esto se debe a un aumento en la eficiencia de detección y cambio de programación para la adquisición de datos del TNS. Por otro lado, se puede observar una tendencia creciente durante estos años, un resultado que concuerda con que el Sol va hacia el mínimo de actividad.

Conclusiones

En este trabajo, se realizó un análisis de estabilidad estadística de las partículas neutras registradas del 2004 al 2016, por el Telescopio de Neutrinos Solares instalado en Sierra Negra, Puebla. Uno de los resultados inmediatos es que se obtuvieron datos limpios, es decir, datos sin influencia de las distintas fuentes que generan errores. Esto ayudará a saber qué datos descartar y en qué intervalo de tiempo se requiere hacer un análisis más detallado si se requiere extraer información física de los mismos y estudios de física solar. Además, al usar software libre (R), nos da libertad de examinar el código, de usarlo sin restricciones, de distribuirlo y modificarlo a nuestra conveniencia; de este modo, cualquiera puede acceder y entender el código utilizado para llevar a cabo el presente trabajo, de esta manera se pueden reproducir los resultados obtenidos y mejorar y adaptar los códigos a análisis estadísticos de datos que se requieran en el futuro, ahorrando tiempo en programación.

Con base en las medidas de centralización y dispersión que se obtuvieron por año, se pudo observar que el 50 % de los datos registrados y confiables se encuentran cerca de la media. Por otra parte, el 25 % de los datos se encuentran cerca del máximo de la variación diurna y el otro 75 % más alejado . También, analizando el coeficiente de variación se pudo observar que en la mayoría de los años (excepto 2005 y 2009) la dispersión de los datos que corresponden a los canales S2_A y S3_A fue menor comparado con los otros canales, presentando una variación menor al 8 %. Además, los datos del canal S4_A se dispersan más comparado con los datos de los otros canales de energía, ya que este canal detecta partículas con mayor energía; por lo tanto, las cuentas registradas son más bajas.

También se ha obtenido el comportamiento de la variación diurna desde el año 2004, lo cual nos da una visión de cómo se ha llevado a cabo el registro de datos a lo largo de estos años. Además, se conoce el porcentaje anual de datos confiables, respecto al cual se concluye que el TNS funciona de forma adecuada y que, con base en la dificultad de mantener un detector a 4580 m s.n.m., el porcentaje de datos estables es muy alta. Se conocen los intervalos temporales donde los datos no son estables y/o se tiene que hacer un análisis más detallado y particular para extraer información relevante. También conocemos las variaciones por canal de energía y podemos detectar que canal tiene problemas de software, hardware o de sistema de adquisición de datos.

Se concluye también que la variación diurna y señal del TNS es estable. Los errores de-

tectados en los canales se pueden atribuir a diferentes fenómenos, como son errores en la electrónica, variaciones de voltaje, fallas en el suministro de energía eléctrica y errores de programación, que pueden generar picos y/o caídas en la toma de datos y ceros.

Finalmente, se observó que el programa de adquisición de datos es susceptible al hardware, ya que el cambio del servidor de adquisición y almacenaje de los datos genera errores de fecha y hora, retrasos en el tiempo de 1 a varios segundos.

El presente análisis concluye que el TNS en Sierra Negra funciona de forma estable y los datos registrados pueden ser usados para análisis y estudios de física solar.

Apéndice A

Fulguraciones solares

Las fulguraciones solares (o llamaradas) son explosiones intensas en el Sol que viene de la liberación de la energía magnética asociada con las manchas solares. La radiación nociva de una llamarada no puede pasar a través de la atmósfera de la Tierra para afectar físicamente a los seres humanos a nivel de tierra. Aunque, cuando es suficientemente intensa, puede afectar la ionosfera de la Tierra e interferir con nuestros sistemas de comunicaciones, como la radio y el GPS, y también perturbar la electrónica de satélites.

Debido la intensidad de emisión en rayos X, las fulguraciones solares se clasifican en A, B, C, M y X. Cada letra representa un aumento de 10 veces en la producción de energía [14]. Así que una X es diez veces una M y 100 veces una C. En las figuras A.1 y A.2 se muestran dos clases de fulguraciones solares.

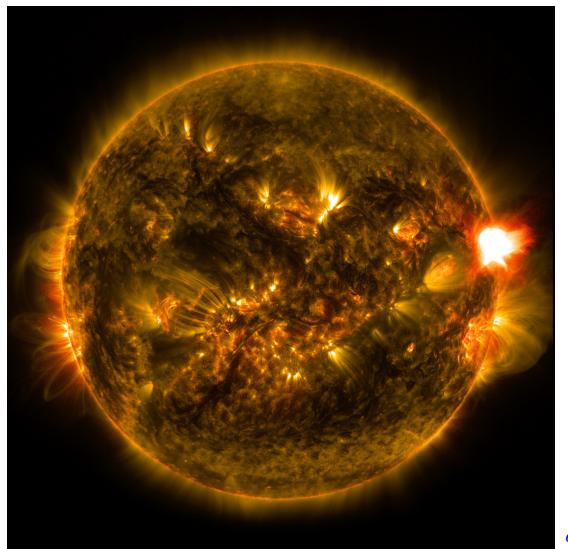


Figura A.1: El Sol emitió una fulguración solar de clase M, alcanzando su punto máximo a las 11:24 p.m. el 12 de enero de 2015. El Observatorio de Dinámica Solar de la NASA, que observa el sol constantemente, capturó una imagen del evento.

^a<https://www.nasa.gov/content/goddard/>

nasa-releases-images-of-1st-notable-solar-flare-of-2015

Se recomienda ver las fotos y animaciones de las fulguraciones solares en la pestaña de galerías que se encuentra accediendo al enlace de cada imagen.

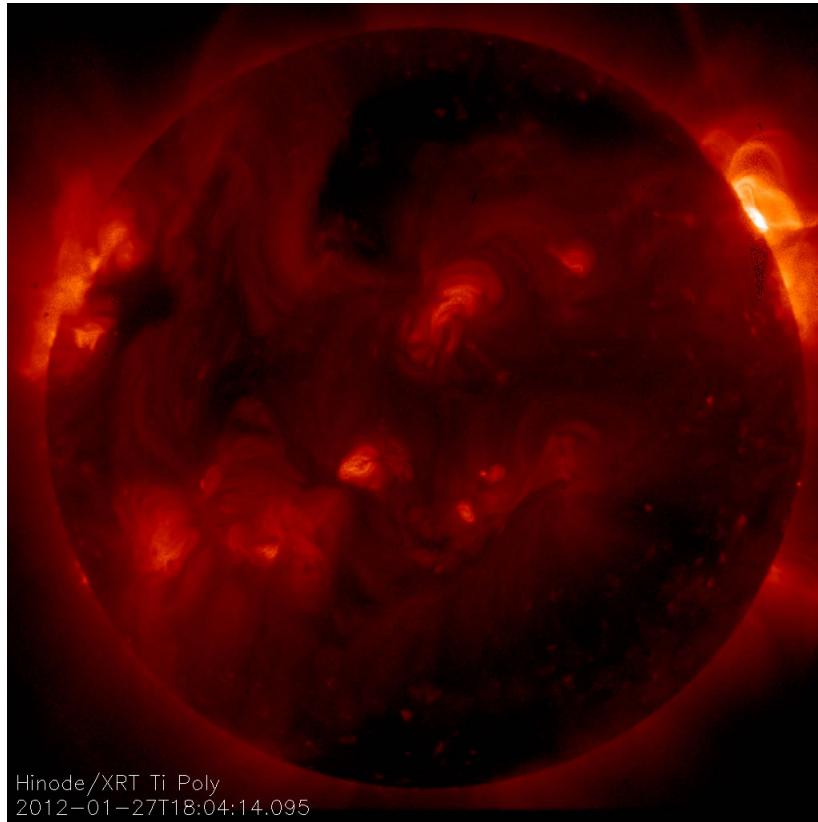


Figura A.2: El 27 de enero de 2012, estalló una fulguración intensa de clase X. Las fulguraciones de clase X son las explosiones más intensas. Aquí se ve una imagen de la llamada capturada por el telescopio de rayos X en Hinode. Esta imagen muestra una emisión de plasma calentado a más de ocho millones de grados durante el proceso de liberación de energía de la llamada. Imagen tomada de https://www.nasa.gov/multimedia/imagegallery/image_feature_2169.html

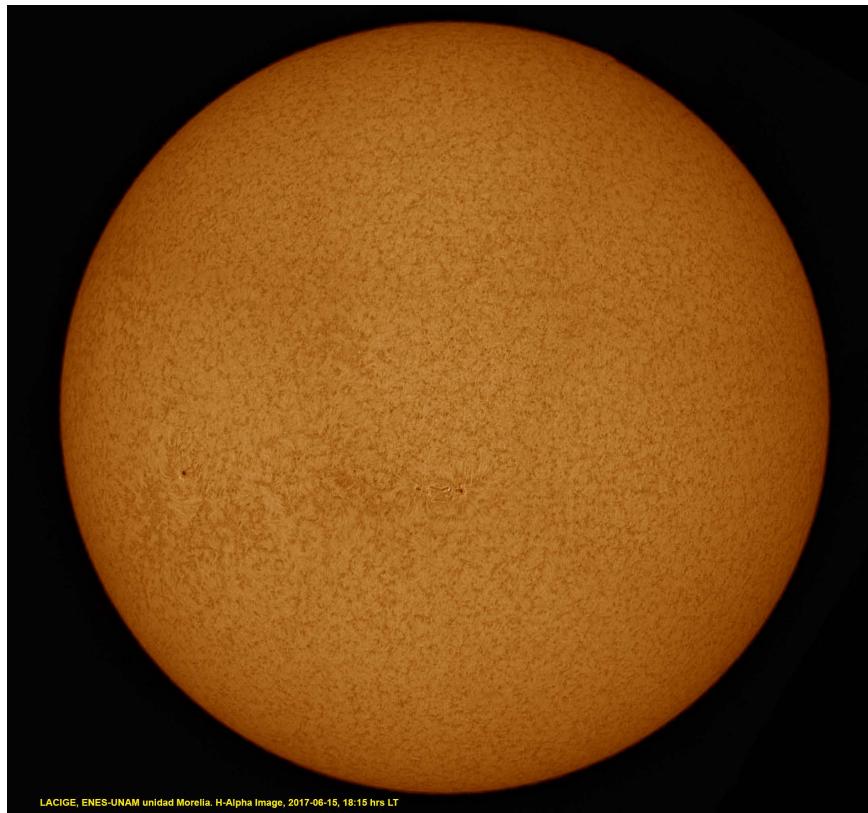


Figura A.3: Se muestra la imagen del Sol en el mínimo de actividad, nótese las diferencias con las figuras A.1 y A.2. Es una imagen en el óptico (H-Alfa), que muestra la fotósfera solar, tomada por LACIGE, ENES-UNAM unidad Morelia.

Referencias

- [1] CARSLAW, D. C., Y ROPKINS, K. openair — an r package for air quality data analysis. *Environmental Modelling and Software* 27–28, 0 (2012), 52–61. [26](#)
- [2] DE LUGO, M. A. B. Cien años de rayos cósmicos. *Rev. R. Acad. Cienc. Exact. Fís. Nat.* (2011). [IX](#)
- [3] GONZÁLEZ, L. X. *El Telescopio de Neutrones Solares en Sierra Negra y Aceleración de Iones en la Atmósfera Solar*. Tesis de Doctorado, Universidad Nacional Autónoma de México, 2006. [IX](#), [IX](#), [X](#), [X](#), [7](#), [11](#)
- [4] GONZÁLEZ, L. X., ET AL. Energy spectrum for the solar neutron event of september 7, 2005, derived from the snt at sierra negra. In *30th ICRC* (2008). [X](#)
- [5] GRIEDER, P. K. F. *Cosmic Rays at Earth*, 1st edition. 27th july 2001 ed. Elsevier Science. [3](#), [4](#), [5](#)
- [6] GROLEMUND, G., Y WICKHAM, H. Dates and times made easy with lubridate. *Journal of Statistical Software* 40, 3 (2011), 1–25. [25](#)
- [7] HAYANON. What are cosmic rays? *Solar-Terrestrial Environment Laboratory, Nagoya University* (2006).
- [8] HUA, X.-M., Y LINGENFELTER, R. Solar flare neutron production and the angular dependence of the capture gamma-ray emission. *Solar Physics* 107, 2 (1987), 351–383. [X](#)
- [9] INSTITUTO DE GEOFÍSICA, UNAM. www.geofisica.unam.mx/observatorios/rayos_cosmicos/grupo_raycos/antecedentes.html, 09 de Marzo de 2017. [X](#), [X](#), [10](#)
- [10] KAHLE, D., Y WICKHAM, H. ggmap: Spatial visualization with ggplot2. *The R Journal* 5, 1 (2013), 144–161. [62](#)
- [11] MENDOZA-ORTEGA, B., ET AL. Chismes espaciales, 2001. [8](#)
- [12] MURAKI, Y., ET AL. Solar Flare and Neutron Telescope. *Nuclear Physics B.* (1998). [7](#)
- [13] MURAKI, Y., ET AL. Simultaneous observation of solar neutrons from the iss and high mountain observatories in association with a flare on july 8, 2014. [X](#)

- [14] NASA. https://www.nasa.gov/mission_pages/sunearth/news/X-class-flares.html, 20 de Marzo de 2017. [59](#)
- [15] OTAOLA, J. A., Y VALDÉS-GALICIA., J. F. *Los rayos cósmicos: mensajeros de las estrellas*. Fondo de Cultura Económica, 1992. [IX](#), [4](#), [5](#), [8](#)
- [16] PARADIS, E. R para principiantes. [17](#), [18](#), [19](#)
- [17] PLIEGO, F. F. *Lenguaje R aplicado al análisis de Calidad de Aire*. Environmental Risk Management, Septiembre de 2012. [26](#)
- [18] ROSS-IHAKA, Y ROBERT-GENTLEMAN. R: a language for data analysis and graphics. *Journal of Computational and Graphical Statistics* 5: 299–314 (1996). [17](#)
- [19] ROSSI, B. B. *Cosmic Rays*. McGraw-Hill, 1964. [IX](#)
- [20] SANTANA, J. S., Y FARFÁN., E. M. *El arte de programar en R: un lenguaje para la estadística*, 1 ed. Instituto Mexicano de Tecnología del Agua. [19](#)
- [21] SCHLAEPPER, H. Cosmic Rays. *The Century of Space Science*. (2003). [IX](#)
- [22] TODOR STANEV. *High Energy Cosmic Rays*, second edition ed. Praxing Publishing. [3](#)
- [23] TSUCHIYA, H., ET AL. Detection efficiency of a new type of solar neutron detector calibrated by an accelerator neutron beam. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 463, 1–2 (2001), 183 – 193. [7](#), [9](#)
- [24] VALDÉS-GALICIA, J., ET AL. Solar neutron events as a tool to study particle acceleration at the sun. *Advances in Space Research* 43, 4 (2009), 565 – 572. [7](#)
- [25] VALDÉS-GALICIA, J. F., ET AL. A New Solar Neutron Telescope in Mexico. In *28th International Cosmic Ray Conference (ICRC)* (2004). [X](#), [7](#)
- [26] WATANABE, K., ET AL. Solar neutron events of october-november 2003. [IX](#)
- [27] WICKHAM, H. The split-apply-combine strategy for data analysis. *Journal of Statistical Software* 40, 1 (2011), 1–29.
- [28] WICKHAM, H. Tidy data. *The Journal of Statistical Software* 59 (2014). [21](#)
- [29] WICKHAM, H., FRANCOIS, R., HENRY, L., Y MÜLLER, K. *dplyr: A Grammar of Data Manipulation*, 2017. R package version 0.7.1. [25](#)