



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, CDMX, 2018



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

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

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

Hoja de Datos del Jurado

1. Datos de alumno

Barón
Martínez
Griselda
5548045380
Universidad Nacional Autónoma de
México
Facultad de Ciencias
Matemáticas
412053104

2. Datos del tutor

Dr.
Luis Xavier
González
Méndez

3. Datos del sinodal 1

Dr.
José Francisco
Valdés
Galicia

4. Datos del sinodal 2

Dr.
Federico Juan
Sabina
Císcar

5. Datos del sinodal 3

Dra.
Blanca Emma
Mendoza
Ortega

6. Datos del sinodal 4

Act.
Jaime
Vázquez
Alamilla

7. Datos del trabajo escrito

Análisis de estabilidad de partículas neutras detectadas por el Telescopio de
Neutrones Solares en Sierra Negra, Puebla.
57 p
2018

*Con todo cariño y amor a mis padres y hermanos.
A la Facultad de Ciencias y a la Universidad Nacional Autónoma de México,
por la formación que me han dado.
Es gracias a ustedes que es posible el presente trabajo.*

Agradecimientos

Antes que nada quiero agradecer a mis padres y hermanos por el apoyo que me han brindado, especialmente a mi hermana Ana que siempre ha estado conmigo de manera incondicional.

Al Dr. Luis Xavier González Méndez por haberme dado la oportunidad de participar en el proyecto, por la asesoría, el apoyo, la confianza y sus conocimientos.

Quiero expresar también mi más sincero agradecimiento al Dr. José Francisco Valdés Galicia por los datos, el apoyo y las ideas. A todos y cada uno de mis sinodales, gracias por el tiempo dedicado a este trabajo y por sus valiosas sugerencias.

Al PUIC, Programa Universitario de Estudios de la Diversidad Cultural y la Interculturalidad, por el apoyo brindado durante mis estudios y en la realización de esta tesis.

A todos los que contribuyen al proyecto R y por su filosofía de software libre.

Gracias Universidad Nacional Autónoma de México y querida Facultad de Ciencias por la excelente educación y por hacer de esta travesía la más agradable. Gracias a mis amigos y a todas la personas que, ya sea de manera directa o indirecta, han sido partícipes de este proceso.

Índice general

Resumen	XI
Introducción	XIII
1. Los rayos cósmicos	1
1.1. Composición	2
1.2. Espectro de energía	2
1.3. Modulación solar	4
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 orden de los datos	21
3.3. Paquetes usados de R	24
3.3.1. dplyr	25
3.3.2. lubridate	25
3.3.3. Openair	27
4. Análisis de estabilidad	29
4.1. Metodología	30
4.2. Análisis de los datos	32
4.3. Señal total del TNS	48
4.4. Resultados	50
Conclusiones	51
A. Imágenes de Actividad Solar	53
Referencias	55

Índice de figuras

1.1. Espectro de energía de los rayos cósmicos.	3
1.2. Observaciones mensuales del número manchas solares	5
1.3. Decrecimiento Forbush detectado por el monitor de neutrones de la Ciudad de México el 22 de junio de 2015.	5
1.4. Esquema de la variación diurna, tomada de [1].	6
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.	12
3.1. Entrada de datos en R	19
3.2. Creando listas y dataframes.	20
4.1. Comportamiento de datos para el año 2005.	32
4.2. Datos normalizados a la media para 2005.	33
4.3. Incrementos de datos horarios en 2005 (marzo-septiembre).	34
4.4. Comportamiento de datos para el año 2009.	35
4.5. Datos normalizados a la media para 2009.	35
4.6. Incrementos de datos horarios en 2009.	36
4.7. Comportamiento de datos para el año 2011.	37
4.8. Datos normalizados a la media para 2011	37
4.9. Incrementos de datos horarios en 2011.	38
4.10. Variación diurna observada en agosto y septiembre de 2011.	38
4.11. Comportamiento de datos para el año 2012.	39
4.12. Datos normalizados a la media para 2012.	40
4.13. Incrementos de datos horarios en 2012.	40
4.14. Comportamiento de datos para el año 2013.	41
4.15. Datos normalizados a la media para el año 2013.	42
4.16. Incrementos de los datos horarios en 2013.	42
4.17. Comportamiento de datos para el año 2014.	43
4.18. Datos normalizados a la media para el año 2014.	44
4.19. Incrementos de los datos horarios en 2014.	44
4.20. Comportamiento de datos para el año 2015.	45
4.21. Datos normalizados a la media para el año 2015.	46
4.22. Datos horarios normalizados a la media en 2015 (Febrero - Marzo).	46

4.23. Datos horarios normalizados a la media en 2015 (Abril - Mayo).	47
4.24. Serie diferenciada de datos horarios en 2015.	47
4.25. Señal completa para el canal $S1_A$ con una razón de conteo de 24 horas.	48
4.26. Máximos de la variación diurna.	49
A.1. Fulguración solar de clase M, 12 de enero de 2015	53
A.2. Fulguración solar de clase X, 27 de enero de 2012	54
A.3. El Sol con baja actividad	54

Índice de tablas

1.1. Composición química de los rayos cósmicos primarios [2].	2
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.	13
2.3. Datos del 1 de junio de 2008, razón de conteo 10 segundos.	14
2.4. Datos del 10 de marzo de 2008, razón de conteo 3 segundos.	15
3.1. Errores evidentes en los datos.	26
4.1. Resultados del análisis estadístico para 2005	33
4.2. Resultados del análisis estadístico para los datos del año 2009.	34
4.3. Resultados del análisis estadístico para los datos del año 2011.	36
4.4. Resultados del análisis estadístico para los datos del año 2012.	39
4.5. Resultados del análisis estadístico para los datos del año 2013.	41
4.6. Resultados del análisis estadístico para los datos del año 2014.	43
4.7. Resultados del análisis estadístico para los datos del año 2015.	45

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 estos canales pertenecen 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 depositada y la 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 ordenados se procede con el análisis gráfico y estadístico para después pasar a la interpretación de los resultados obtenidos. Se muestran las variaciones producidas 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. De esta manera, el análisis nos permite asegurar que se está trabajando con datos confiables para realizar estudios básicos de física solar, gracias a que existen elementos para poder conocer y diferenciar las afectaciones de los fenómenos puramente eléctrico-electrónicos y/o de otra índole.

Como los datos registrados no son series de tiempo estacionarias, se realizó una transformación para volverlas estacionarias o, al menos, para que varíen alrededor de algún valor. Esto con el fin de poder aplicar gráficas de control estadístico con líneas de control $\mu \pm 3\sigma$ y corroborar que los incrementos sean estables.

Finalmente, se obtuvieron datos limpios, es decir, datos sin influencia de las distintas fuentes que generan errores. Se encontró que el canal $S3_A$ es el que ha presentado más fallas electrónicas y que el $S4_A$ tiene variaciones más altas con respecto a los tres primeros canales de energía.

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

También se obtuvo la señal total del detector para tener un panorama general del comportamiento del TNS y compararla con el ciclo de manchas solares.

Además, se pudo observar que el máximo de la variación diurna se encuentra, en su mayoría, entre las 12 y 18 horas, lo que es consistente con la variación diurna en los observatorios a nivel de tierra. Ha detectado dos eventos de neutrones solares y tiene un registro del 86 % de datos desde su puesta en marcha en julio de 2004. De esta manera, podemos concluir que el detector funciona de forma adecuada en términos generales.

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 antiprotón 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 [3].

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. Se producen principalmente en estrellas, remanentes de supernovas y núcleos activos de galaxias. La composición de los rayos cósmicos es de aproximadamente 93 % de protones, 6.3 % de partículas alfa y 0.7 % son núcleos de elementos más pesados. Hay tres tipos de rayos cósmicos: los galácticos y extragalácticos, anómalos y solares. Los *rayos cósmicos galácticos y extragalácticos* dominan el espectro de energía desde 10^6 eV hasta 10^{21} eV² (ver Figura 1.1). Los *rayos cósmicos anómalos* se originan en el medio interplanetario y se denominan así porque su composición es inusual; no siguen las abundancias naturales predichas para los diferentes isótopos³ [5]. En la atmósfera del Sol también se producen una gran cantidad de partículas a las que llamamos *rayos cósmicos solares* y cuya energía domina el espectro desde 10^3 eV hasta 10^6 eV [6]. 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 emite en todo el espectro electromagnético y puede acelerar partículas hasta 10^{10} eV, la mayor parte son partículas cargadas (electrones, protones y núcleos de elementos más pesados), éstas son desviadas por los 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 los campos magnéticos, conservan información relevante para comprender y calcular el espectro de los iones y su tiempo de producción [7].

¹Galardonado con el Premio Nobel de Física en 1936.

²Los 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 [2] y [4].

³Núcleos del mismo elemento que tienen diferente número de neutrones.

Los neutrones solares contienen información importante sobre los mecanismos de aceleración de iones en la atmósfera del Sol [8]. Con base en que el tiempo de vida media de los neutrones libres es de 886 segundos, tienen que ser de muy alta energía (relativistas) para llegar a la Tierra y ser detectados.

El mecanismo de aceleración de los electrones ha sido estudiado mediante las observaciones con rayos X y rayos gamma (γ), mientras que el de los iones aún no se comprende totalmente. 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 [9]. 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 [6].

En el instituto de Geofísica de la UNAM se detectan rayos cósmicos con energías desde los 8.2 GeV¹ con un instrumento que se conoce como Monitor de Neutrones² (MN) [10]. 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 arribo 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 que depositan en el detector y puede diferenciar entre neutrones y protones. El nombre de telescopio se debe a que además de la capacidad mencionada anteriormente, mide la dirección de arribo de las partículas incidentes. Existe una red mundial de TNS instalados en altas montañas (Armenia, Bolivia, China, Hawai, 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[11]. Ha estado funcionando de manera casi continua desde julio de 2004 y ha registrado dos eventos de neutrones solares producidos en las fulguraciones del 7 de septiembre de 2005 y del 8 de julio de 2014, que también fueron detectados por el TNS instalado en Chacaltaya, Bolivia [12, 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 para asegurar que son viables para estudios de física solar. El análisis se realiza para los datos detectados por los cuatro canales que registran partículas neutras, con base en el uso del software estadístico R. El objetivo es localizar los posibles errores generados durante el registro para descartarlos en estudios posteriores, así como encontrar variaciones importantes.

¹Rigidez umbral. Véase el *capítulo 1*

²El MN instalado en C.U. es considerado uno de los detectores de neutrones más estables del mundo [10].

Para saber si las partículas neutras registradas por el TNS son estadísticamente estables, es necesario estudiar su fluctuación a través del tiempo y distinguir las variaciones atribuidas a emisiones solares de las variaciones atribuidas a fallas en el suministro de energía eléctrica, fallas de algunos componentes electrónicos o errores en el sistema de adquisición de datos.

Este trabajo está estructurado en 4 capítulos. En el primer capítulo se explican las generalidades de los rayos cósmicos, como su composición, espectro de energía y las variaciones atribuidas a las emisiones solares.

En el capítulo 2 se muestra la red mundial de Telescopios de Neutrones Solares, una descripción a grandes rasgos 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 este telescopio, ya que durante los 13 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.

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

Los rayos cósmicos

¿Qué son los rayos cósmicos? La respuesta a esta pregunta es, en principio, sencilla. Los rayos cósmicos son partículas de alta energía que llegan a la Tierra procedentes del espacio exterior.

El descubrimiento de la radioactividad por Henri Becquerel en 1896 y los estudios, entre otros, por Marie y Pierre Curie en Francia y por Ernest Rutherford en Canadá establecieron la existencia de materiales radiactivos. Rutherford sugirió que las distintas variedades de radiación descubiertas podían atravesar la materia e *ionizar* el aire, motivando a investigadores de su entorno, entre ellos a Wilson, a detectar la presencia de materiales radiactivos en las proximidades de sus instrumentos de medida [3].

En 1901, Wilson declaró que la ionización espontánea podría deberse a una radiación penetrante de origen cósmico, pero no llegó a comprobar esta hipótesis. Posteriormente, T. Wulf y W. Kohlhořster desarrollarían instrumentación más avanzada que permitiría realizar medidas más precisas; sin embargo, fue hasta 1912 que Victor Hess concluiría con la existencia de una radiación proveniente del espacio exterior. Durante 1911 y 1912 Hess realizó ascensos en globo llevando consigo varios electroscopios, del tipo que Wulf había usado, alcanzando distintas altitudes y obteniendo mediciones relevantes. Fue el 7 de agosto de 1912 que llegó a alcanzar hasta 5350 m de altitud y se dio cuenta que la cantidad de iones atmosféricos aumentaba a partir de cierta altura y, por lo tanto, la radiación que los originaba debía ser extraterrestre [2].

En los años transcurridos, los investigadores han aprendido mucho acerca de estas partículas altamente energéticas. Hoy en día se sabe que los rayos cósmicos son en su mayoría núcleos atómicos despojados de sus electrones por los procesos de aceleración en la interacción desde la fuente hasta la Tierra [14].

Clasificamos a los rayos cósmicos en: solares (RCS), galácticos y extragalácticos (RCG) y Anómalos (RCA). Los RCS se emiten en la atmósfera del Sol, los más energéticos son acelerados principalmente durante las fulguraciones. Dominan el espectro de energía de $10^3 eV$ hasta $10^6 eV$, aunque algunas fulguraciones pueden acelerar RCS hasta $10^{10} eV$ [15]. Los RCG provienen desde fuera del sistema solar y son acelerados principalmente en remanentes de supernovas y núcleos activos de galaxias. Los RCG dominan el espectro de energía desde $10^6 eV$

hasta $10^{21} eV$ y son la fuente principal del fondo de rayos cósmicos que se detectan en tierra. Los RCA son el producto de átomos neutros provenientes del medio interestelar que al entrar al Sistema Solar son ionizados por fotoionización e intercambio de cargas. Una vez ionizados son asimilados por el viento solar y llevados al choque terminal de éste, donde son acelerados a energías del orden de 10 a 100 MeV. Sólo pueden ser detectados por naves espaciales.

Generalmente a las partículas que inciden en la atmósfera terrestre procedentes del espacio, compuestos principalmente de protones y partículas alfa, se les denominan rayos cósmicos primarios. Al interactuar con un núcleo atmosférico producen una cascada de partículas compuesta por rayos γ , electrones, muones, nucleones, etc., que son llamados rayos cósmicos secundarios [16].

A continuación se dan algunas características de los rayos cósmicos.

1.1. Composición

Uno de los hechos importantes acerca de la radiación cósmica se refiere a su composición química, el 93% son protones, 6.3% partículas alfa y 0.7% son núcleos de elementos más pesados (ver Tabla 1.1).

Grupo	Elemento	Porcentaje de abundancia de los RC
Proton	H	93
Alfa	He	6.3
L	Li, Be, B	0.10
M	C, N, O, F	0.42
H	Ne-K	0.14
VH	Ca-Zn	0.04
VVH	Ga-U	2×10^{-6}

Tabla 1.1: Composición química de los rayos cósmicos primarios [2].

1.2. Espectro de energía

El *espectro de energía* de la radiación cósmica primaria describe cómo están distribuidas con respecto a la energía. En la Figura 1.1 se muestran las mediciones del espectro de energía de las diversas componentes de radiación cósmica. El aplanamiento del espectro hacia las bajas energías refleja la influencia del Sol en la propagación de los rayos cósmicos.

El espacio interplanetario no es un vacío por el cual las partículas de rayos cósmicos puedan propagarse libremente, el viento solar¹ “empuja” a estas partículas hacia fuera del Sistema Solar, mientras que la presencia del campo magnético las “guía” hacia dentro. Estas dos fuerzas en competencia no son uniformes ni constantes en el tiempo, por ello, sus efectos en la intensidad de la radiación cósmica detectada en cualquier punto se manifiestan en forma de variaciones temporales [1].

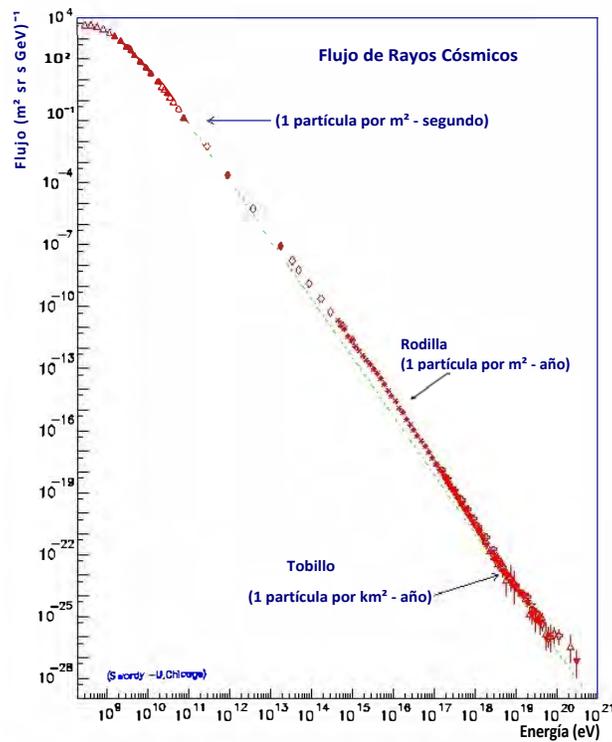


Figura 1.1: Espectro de energía de los rayos cósmicos.

El mecanismo que provoca esta variación secular resulta del balance que se establece entre la convección que genera el viento solar al desplazarse constantemente hacia afuera y la lenta pero segura difusión de las partículas hacia dentro. Como primera consecuencia se tiene que aún en tiempos de mínima actividad el espectro de rayos cósmicos que se observa en la Tierra no es el que llega a las inmediaciones de la heliósfera². La segunda consecuencia de la convección-difusión es que el flujo de rayos cósmicos será progresivamente reducido a medida que éstos se acerquen al Sol [1].

Existe un gran número de mediciones y estudios del sector del espectro con energías menores al “tobillo”; mientras que la existencia de la “rodilla” no está aún totalmente explicada.

¹Plasma generado con la continua expansión de la atmósfera solar.

²Región alrededor del Sol permeada por el plasma solar y su campo magnético. Esta región se extiende hasta, aproximadamente, 122 unidades astronómicas.

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 provienen de fuera de la Galaxia.

1.3. Modulación solar

Durante muchos años se ha observado que la intensidad de los rayos cósmicos medida en la Tierra experimenta variaciones periódicas en función de las emisiones de la actividad del Sol, se dice que estas variaciones son producidas por la *modulación solar* [17]. 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. Se ha observado que 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 1 o 2 años aproximadamente [18]. 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.

Aún se estudian los mecanismos y procesos de generación de los decrecimientos Forbush y se sabe que son un efecto de las líneas del 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 [17]. 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.

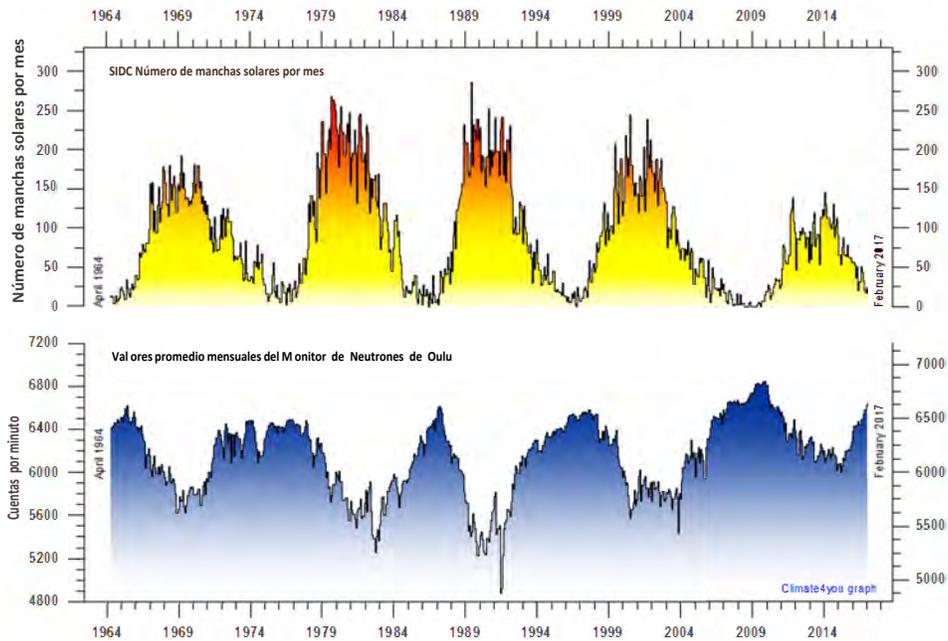


Figura 1.2: Observaciones mensuales del número manchas solares (Solar Influences Data Analysis Center (SIDC)) desde abril de 1964 y medias mensuales de las cuentas del monitor de neutrones de Oulu (Finlandia), ajustados por presión barométrica y eficiencia. Actualización: 1 de marzo de 2017.

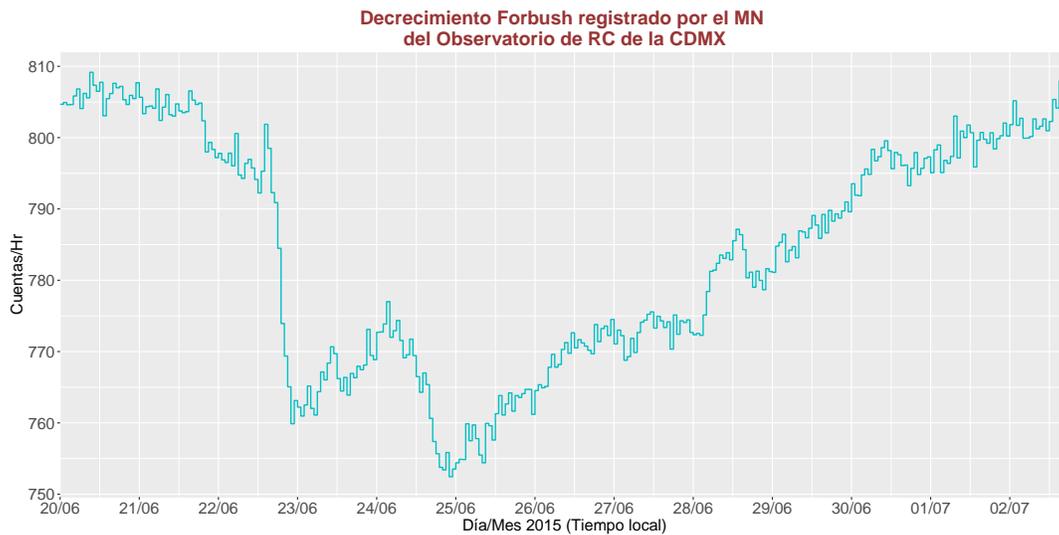


Figura 1.3: Decrecimiento Forbush detectado por el monitor de neutrones de la Ciudad de México el 22 de junio de 2015.

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 [17]. Esta variación es una consecuencia del equilibrio entre el flujo difusivo y convectivo. El flujo neto o resultante del flujo difusivo y el flujo convectivo hace que la intensidad de la radiación cósmica presente un máximo alrededor de las 18 horas en tiempo local y un mínimo sobre las 3 de la mañana (ver Figura 1.4).

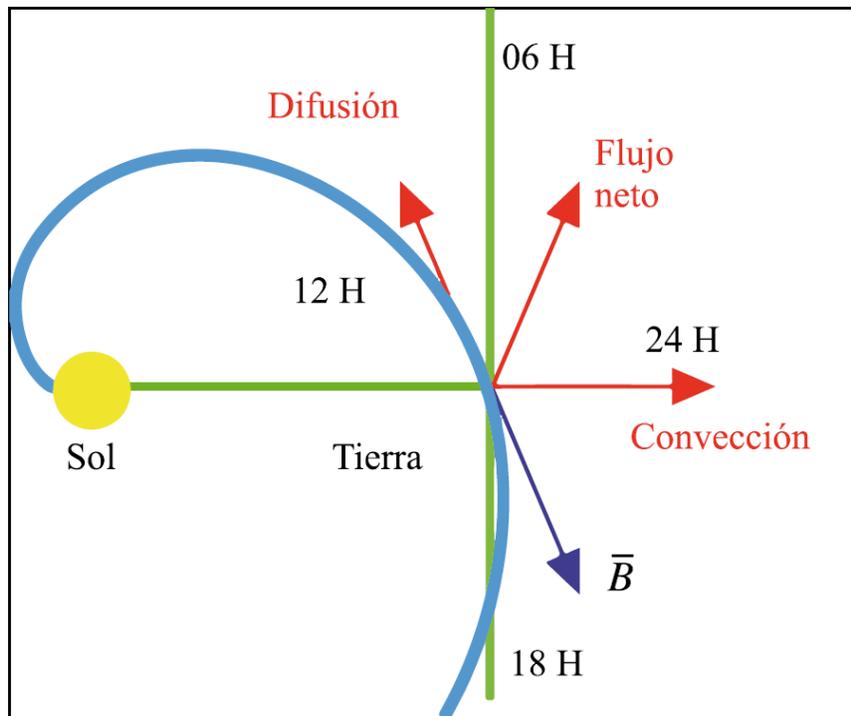


Figura 1.4: Esquema de la variación diurna, tomada de [1].

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 [19]. 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 [20].

Los monitores de neutrones tienen cuentas muy elevadas debido a su alta sensibilidad y no pueden diferenciar la dirección de arribo de las partículas incidentes [21]. Tampoco resuelven la energía de las partículas y no discriminan entre neutrones y protones. Por esta razón, fue 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 arribo de las partículas incidentes [11, 6].

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 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 rigidez umbral va disminuyendo [2]. 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 [22].

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 interactuar 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.

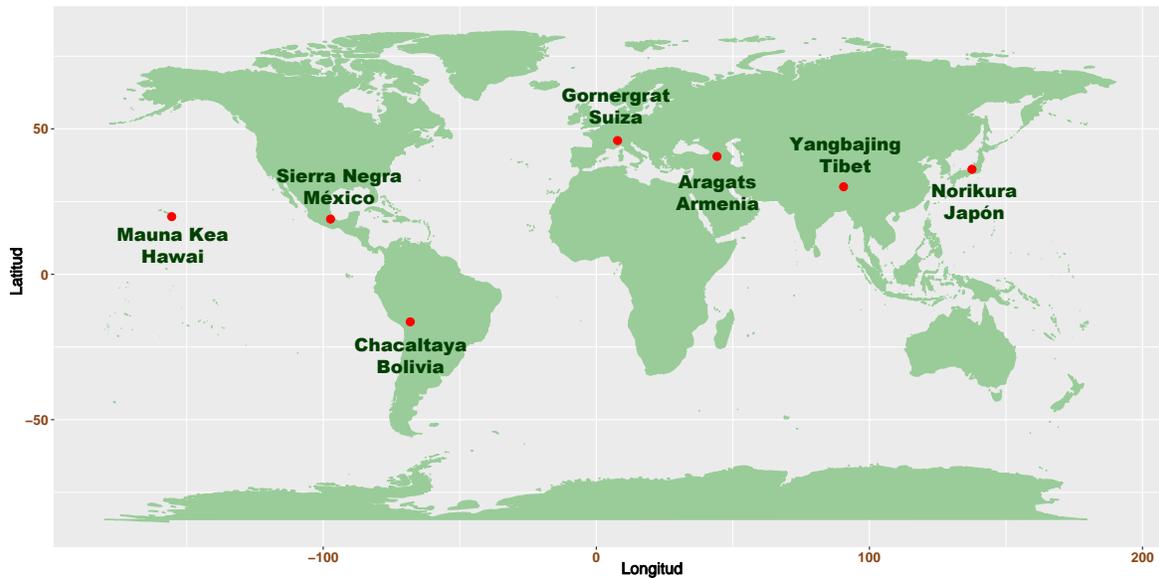


Figura 2.1: Localización geográfica de los TNS alrededor del planeta.

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 el artículo *Detection efficiency of a new type of solar neutron detector calibrated by an accelerator neutron beam* de H. Tsuchiya [21].

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 Hawai y Chacaltaya, lo cual ayuda a tener en observación al Sol de 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 [10], la figura 2.2 muestra un esquema.

¹El Observatorio Solar Mexicano de Gran Altura (OSOMEGA) de la Universidad Nacional Autónoma de México (<http://naolinco.geofisica.unam.mx/observatorios/osmega/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.

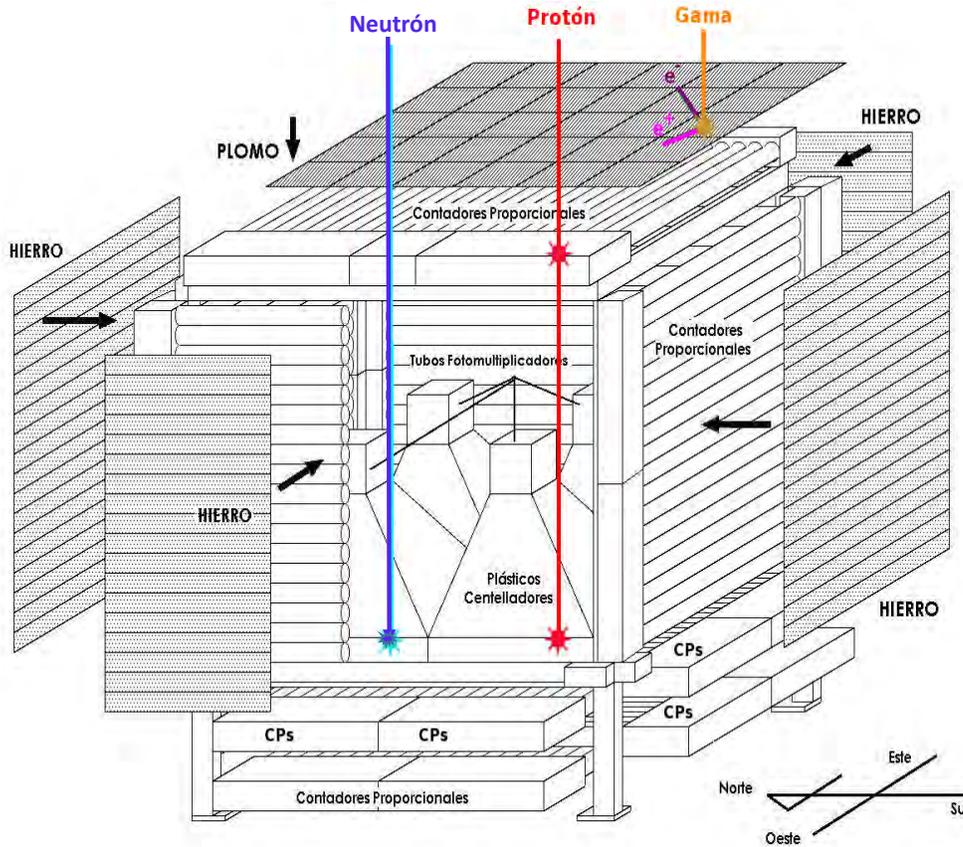


Figura 2.2: Esquema del TNS instalado en la cima del volcán Sierra Negra, Puebla.

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 $4m^2$. Cuando una partícula de la cascada atraviesa uno de estos plásticos una pequeña parte de su energía se emplea en excitar algunas de sus moléculas que inmediatamente se desexcitan emitiendo un breve flash de luz.

La energía depositada por las partículas incidentes se mide en un tubo fotomultiplicador, que es capaz de producir una señal eléctrica de muy corta duración, 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*).

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.

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 arriba se miden usando cuatro capas de CP alineados ortogonalmente debajo de los PC, dos para determinar la dirección de arriba 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 interactuar con los PC. Estos protones secundarios se desvían menos de 13.48° de su dirección original con lo que se asegura que la dirección de arriba del neutrón puede ser determinada [6].

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 arriba 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 calculó 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 [6] 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. Después 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 (*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.

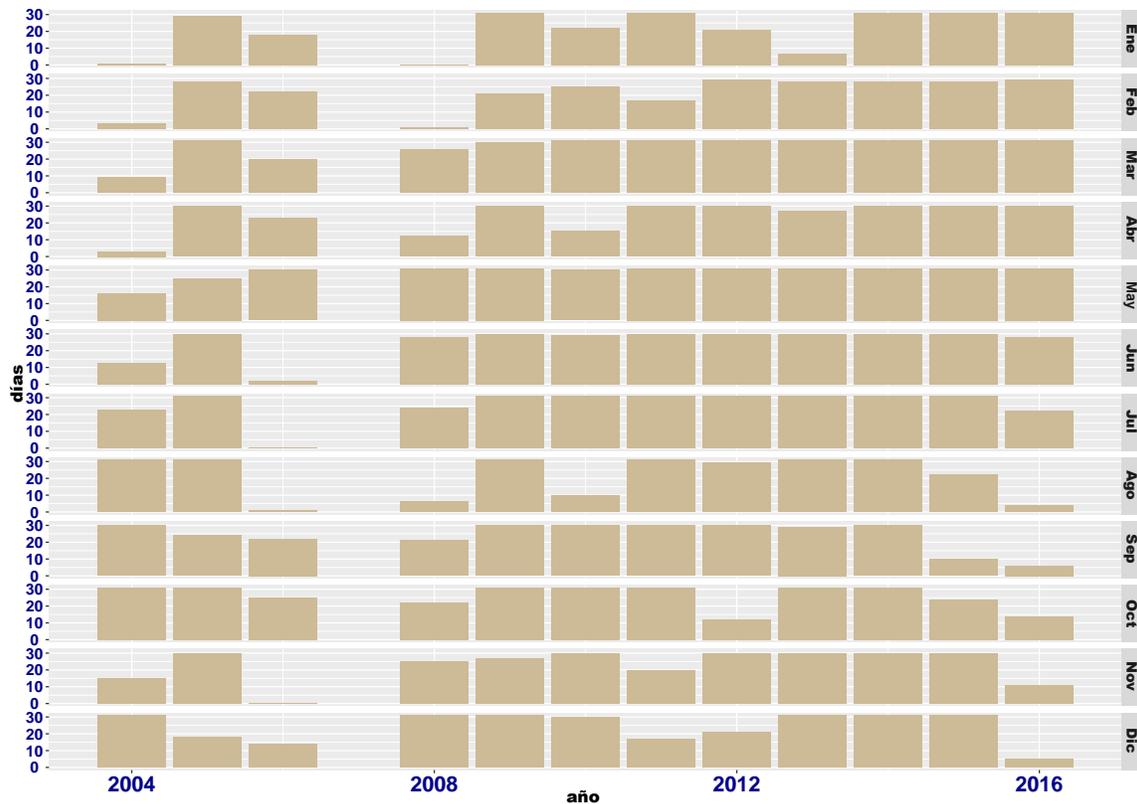


Figura 2.3: Gráfica que muestra el número de días registrados por mes y año.

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, 0720 indica que es el día 20 del mes 07 (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	0720	0	100	8761	3407	1196	541	27371	12855	5036	...
2	2004	0720	1	0	8602	3355	1187	539	27619	13003	5083	...
3	2004	0720	2	0	8705	3402	1261	580	27804	13205	5194	...
4	2004	0720	3	0	8662	3400	1235	565	27645	12987	5051	...
5	2004	0720	4	0	8712	3323	1178	551	27702	12876	4994	...
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	
1437	2004	0720	2356	0	8652	3358	1233	561	27647	12928	5083	...
1438	2004	0720	2357	0	8665	3400	1231	545	27596	12999	5039	...
1439	2004	0720	2358	0	8732	3399	1226	548	27675	12983	5060	...
1440	2004	0720	2359	0	8681	3355	1152	557	27789	12983	5009	...
1441	2004	0721	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 nos dan información interna del sistema de adquisición de datos. Sin embargo, no se utilizarán al momento de procesar 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	...
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	
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 ordenación de datos, en el capítulo 3 se expone 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.

R y desarrollo de códigos

R es un lenguaje de programación para análisis estadístico y gráfico creado por Ross Ihaka y Robert Gentleman [23], 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 [24].

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 esta disponible en varias formas: el código fuente escrito principalmente en C (y algunas rutinas en Fortran), esencialmente para maquinas 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 muy buena 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 [24]. 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 [24].

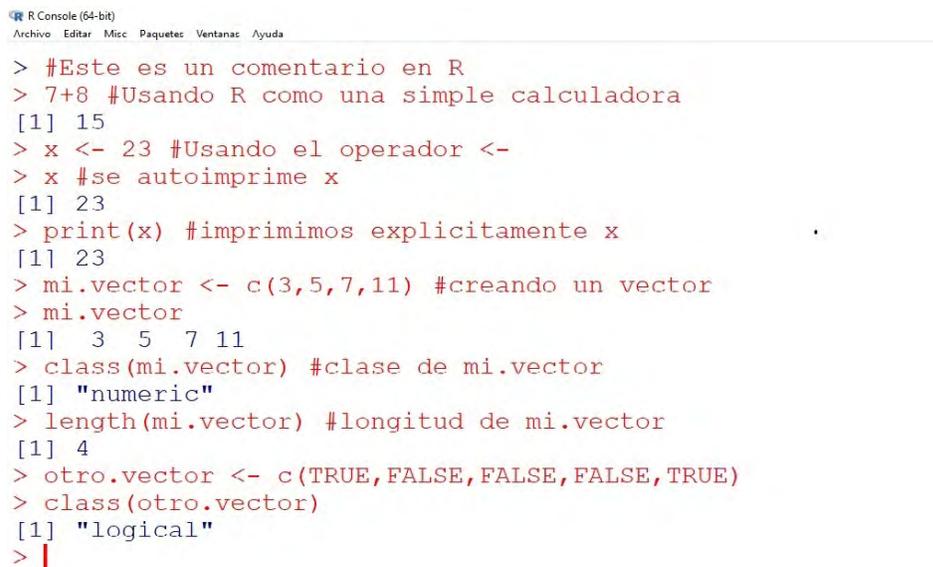
Hay diferentes tipos o clases de datos en R:

- *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) y 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 [25]. 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 explícitamente 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 orden de los 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 [26] 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* [27]. 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 Golemund 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 Golemund y Hadley Wickham [28].

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.

Como el TNS registra en tiempo local, es necesario convertir las fechas y horas primero en tiempo universal y despues 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 ese periodo. 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.

Lo primero que se hizo fue comparar los datos con y sin anticoincidencia para cada canal de energía. En este paso 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 como en la Tabla 3.1.

date	S1_A	S2_A	S3_A	S4_A
2009-02-21 15:20:10	16777215	16777215	16777215	16777215
2009-02-21 15:20:20	16777215	16777215	16777215	16777215
2009-02-21 15:20:30	16777215	16777215	16777215	16777215
2009-02-21 15:20:40	16777215	16777215	16777215	16777215
2009-02-21 15:20:50	16777215	16777215	16777215	16777215
2009-02-21 15:21:00	16777215	16777215	16777215	16777215
2009-02-21 15:21:10	16777215	16777215	16777215	16777215
2009-02-21 15:21:20	16777215	16777215	16777215	16777215
2009-02-21 15:21:30	16777215	16777215	16777215	16777215
2009-02-21 15:21:40	16777215	16777215	16777215	16777215
2009-02-21 15:21:50	16777215	16777215	16777215	16777215

Tabla 3.1: Errores evidentes en los datos.

También era fácil ver que se trataba de errores aquellos datos individuales 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 relevantes. Para esto, fue de mucha utilidad el paquete `Openair`.

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* [29] y en *Lenguaje R aplicado al análisis de Calidad de Aire* [30].

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 se alcanzó el objetivo de tener datos ordenados y que la fecha y hora de éstos sea reconocida como tal en R, además se limpiaron 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. Un ejemplo son los datos de 2008, hasta que se graficaron se observó su comportamiento. Véase el siguiente capítulo para las gráficas de variación por año.

Análisis de estabilidad

A lo largo de 13 años, el TNS ha estado registrando la radiación cósmica y las variaciones causadas por las emisiones solares. Debido a la importancia que presenta el uso correcto de esta información, es conveniente realizar el análisis exploratorio con el fin de determinar cambios en la serie de datos que corresponden a 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. Para que los datos del TNS sean estables, su variación cuando no hay eventos y/o emisiones solares significativas no debe mostrar observaciones inusuales. De acuerdo con la gráfica 2.3, se pudo observar que hay años donde se registraron una cantidad insuficiente de datos, por lo que no se podría hablar de una estabilidad. Por ello el análisis se hará para los años 2005, 2009 y del 2011 al 2015.

Es importante hacer notar que los datos del TNS, al ser de conteo de partículas, son datos discretos. Estos datos están asociados a variables que toman un conjunto finito o numerable de valores posibles. Dentro de las distribuciones discretas, la que mejor modela este tipo de datos es la distribución de Poisson, ya que representa el número de ocurrencias por unidad de tiempo. Sin embargo, para estudiar el comportamiento de un conjunto de variables aleatorias que está indexado en el tiempo, se necesita un concepto que abarque situaciones más generales que las sucesiones de variables aleatorias; por esta razón se introducen las siguientes definiciones.

Definición 4.0.1. *Un proceso estocástico es una colección de variables aleatorias $(X_t)_{t \in T}$ indexadas en un conjunto T y definidas en algún espacio de probabilidad $(\Omega, \mathcal{F}, \mathbb{P})$.*

Interpretamos al conjunto de índices T como un parámetro temporal y al proceso estocástico como la evolución temporal de algún fenómeno que se rige por el azar. Dada ω , $X(\omega, t)$, como una función de t , se denomina realización. Por lo tanto, una serie de tiempo es una realización de ciertos procesos estocásticos [31].

Definición 4.0.2. (Incrementos independientes) *Se dice que un proceso estocástico tiene sus incrementos independientes si para $t_1 < t_2 < \dots < t_n$, las variables*

$$X_{t_2} - X_{t_1}, X_{t_3} - X_{t_2}, \dots, X_{t_n} - X_{t_{n-1}}$$

son independientes.

Una serie de tiempo se dice que es *estacionaria* si el desarrollo de ésta no depende del momento en el cual se empezó a observar, es decir, distintos intervalos de la serie se verán aproximadamente igual en otros intervalos de la misma longitud [32]. Mas adelante se verá que la serie de tiempo de partículas neutras detectadas por el TNS no tiene un comportamiento estacionario, pero sí tiene incrementos independientes, ya que la cantidad en cada intervalo de tiempo no depende de la cantidad de otro intervalo.

4.1. Metodología

Al comprobar la estabilidad de las observaciones en el tiempo, los cálculos para la media muestral y la varianza muestral tratan todas las observaciones por igual. Se supone que no hay tendencias aparentes en los datos a lo largo del tiempo y no hay observaciones inusuales. Otra forma de decir esto, es que el proceso que produce las observaciones está en *control estadístico*. El concepto del control estadístico permite la variabilidad en las observaciones pero requiere que el patrón de variabilidad sea el mismo a lo largo del tiempo [33].

El análisis de la serie de datos de partículas neutras consiste en detectar por medio gráfico y cuantitativo la existencia de cambios anormales. El esquema metodológico del análisis exploratorio se realiza por año para los cuatro canales. Empieza con el estudio de un gráfico de la señal del TNS en ese año y muestra el porcentaje de datos que se tienen por mes con la resolución mínima de conteo. Además, en esta gráfica se puede observar el histograma, una distribución de frecuencias donde los rectángulos representan frecuencias relativas .

Una vez que conocemos los datos con la razón de conteo registrada, se comparan las señales de los cuatro canales de energía mediante un gráfico con la serie de tiempo diaria y normalizada con respecto a la media; también se muestran las medidas de tendencia central y de dispersión. Antes de normalizar se obtuvieron los datos con razón de conteo un día, cuando el día no contiene todos los datos con razón de conteo mínima, entonces se toma como NA, de lo contrario, se suman los datos a la tasa de conteo requerida.

Por último, se lleva a cabo la prueba de control estadístico para los incrementos de los datos horarios, el resultado se mostrará de manera gráfica. De acuerdo con la regla empírica, si el proceso está en control estadístico, solo alrededor del 5% de las observaciones quedarán fuera de los límites de control $\mu + 3\sigma$ y $\mu - 3\sigma$. Estos límites nos permiten distinguir entre variación típica y variación que es especialmente grande y podría deberse a causas especiales o asignables.

Antes de pasar a los resultados del análisis se definen las medidas de tendencia central y dispersión.

La media muestral de un conjunto de n mediciones $x_1, x_2, x_3, \dots, x_n$ es la suma de éstas dividido por n . La media muestral se denota por \bar{x}

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} \quad (4.1)$$

La mediana muestral de un conjunto de n mediciones, $x_1, x_2, x_3, \dots, x_n$, es el valor que queda en medio cuando los datos se ordenan de manera creciente. En general, es el valor que divide los datos a la mitad, es decir, el 50 % queda por arriba de la mediana y el otro 50 % por abajo.

Como el número de observaciones es bastante grande, a veces es útil dividir los datos ordenados en cuartos para ampliar la noción de mediana. Así como el punto de división en mitades se llama mediana, los puntos para la división en cuartos se llaman *cuartiles*. De manera que el primer cuartil deja por debajo al 25 % de las observaciones y 75 % por arriba, el segundo cuartil coincide con la mediana, mientras que el tercer cuartil deja por debajo al 75 % de los datos y 25 % por arriba.

Las medidas de tendencia central reducen la información de la muestra a un solo valor. Sin embargo, este valor central será mas o menos representativo de dicha muestra dependiendo de la dispersión que las medidas individuales tengan respecto a dicho centro. Para analizar la representatividad de las medidas de centralización se definen las llamadas medidas de dispersión. Estas nos indicarán la variabilidad de las observaciones con respecto a su valor promedio, es decir, si se encuentran muy o poco esparcidos en torno a su centro.

Para obtener una medida de dispersión, se deben eliminar los signos de las desviaciones antes de promediar. Una forma de eliminar la interferencia de los signos es tomando el cuadrado de los números. Se construye una medida de dispersión, llamada *varianza muestral* (s^2), agregando las desviaciones cuadradas y dividiendo el total por el número de observaciones menos uno.

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1} \quad (4.2)$$

Para obtener una medida de la variabilidad en la misma unidad, tomamos la raíz cuadrada positiva de la varianza, llamada *desviación estándar*. La desviación estándar en lugar de la varianza sirve como una medida básica de variabilidad.

$$s = \sqrt{s^2} = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}} \quad (4.3)$$

4.2. Análisis de los datos

A continuación se muestran las gráficas obtenidas, empezando con 2005 y continuando conforme avanzaron los años .

La figura 4.1 muestra una síntesis de los datos registrados con la resolución mínima de conteo para cada canal de energía. La barra color gris representa los datos existentes durante todo el año, mientras que el sombreado rosa muestra los datos faltantes. De esta manera, para cada mes se sabe el porcentaje de datos considerados para el análisis, este porcentaje se muestra en la parte superior de la señal del detector durante el año. Por ejemplo, para enero hay 62.2% de datos, donde los que faltan se encuentran, la mayoría, a fin de mes. Vemos también que enero, febrero y diciembre son los meses en los que se tiene menos del 65% de datos. Además, en la gráfica se muestran las principales estadísticas para esta serie, como el valor mínimo, máximo, media, mediana y el porcentaje total de datos faltantes; con el fin de conocer lo mejor posible a la serie que se va a estudiar.

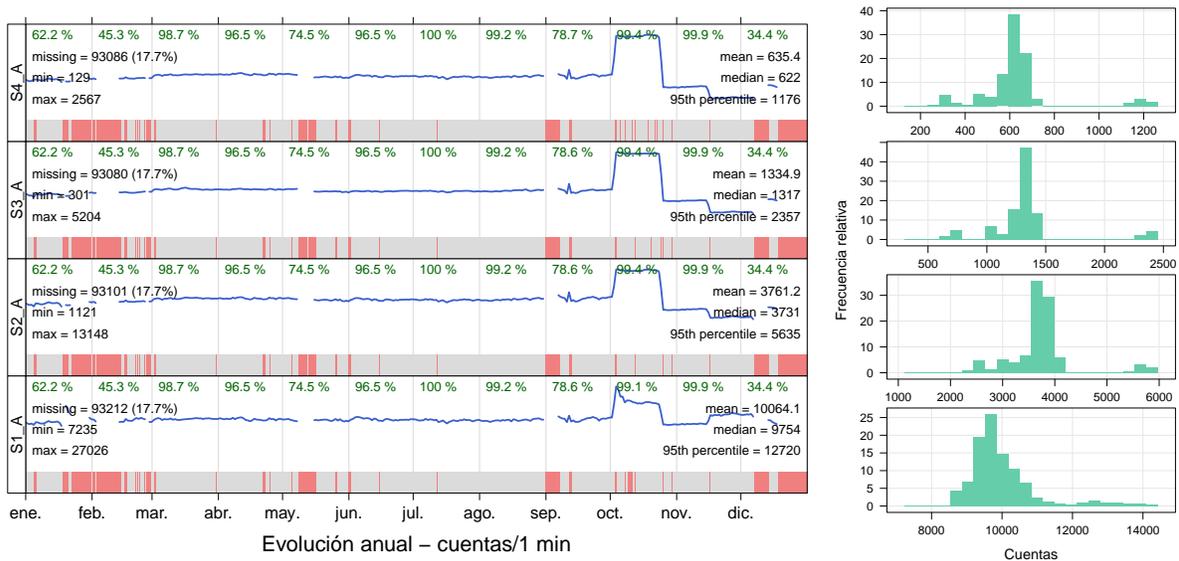


Figura 4.1: Comportamiento de datos para el año 2005.

Al analizar la señal, se observa que a partir de octubre hay una variación importante presente en todos los canales, ésta dura casi todo el mes y en noviembre vuelven a bajar las cuentas. El $S1_A$ ($E \geq 30\text{MeV}$) es el único que sigue un comportamiento diferente después de este mes y se refleja en la distribución de los datos.

	Mínimo	1er cuartil	Mediana	Media	3er cuartil	Máximo	desv.est
S1_A	0.0444	0.9412	0.9699	1	1.0233	1.8910	0.11
S2_A	0.0412	0.9583	0.9938	1	1.0242	1.9687	0.17
S3_A	0.0388	0.9535	0.9881	1	1.0196	2.0220	0.25
S4_A	0.0367	0.9374	0.9815	1	1.0212	2.0795	0.28

Tabla 4.1: Resultados del análisis estadístico para datos horarios y normalizados a la media del año 2005.

Con base en la figura 4.2, se observa que en el mes de octubre las cuentas subieron desde un 20 % hasta aproximadamente 90 %, siendo el primer canal el que menos se aleja de la media. Esto último se puede corroborar en la Tabla 4.1, donde la desviación estándar es de 0.11. La variación en octubre debe ser un error electrónico, ya que no se observa una variación gradual, el aumento es más del 50 % y no está correlacionado con una variación en el flujo de rayos cósmicos detectado en otros TNS de la red mundial. Por esta razón se realiza la prueba de control estadístico sobre los datos de marzo a septiembre.

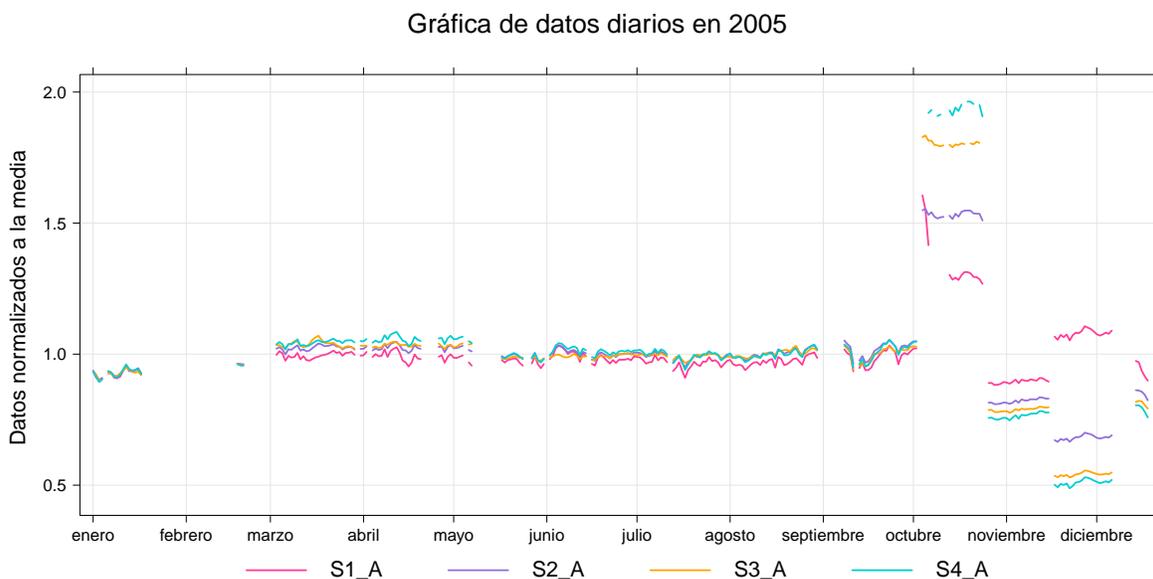


Figura 4.2: Datos normalizados a la media para 2005.

En la figura 4.2 se muestra que la serie de marzo a septiembre no es estacionaria, si aplicáramos la prueba de control estadístico sobre estos datos arrojaría respuestas poco confiables. Lo recomendable es transformar la serie a una estacionaria, para ello tomamos la serie de incrementos. Con base en la figura 4.3, se observa que los datos de marzo a septiembre son estables, ya que los datos fuera de los límites de control es menor al 5 %.

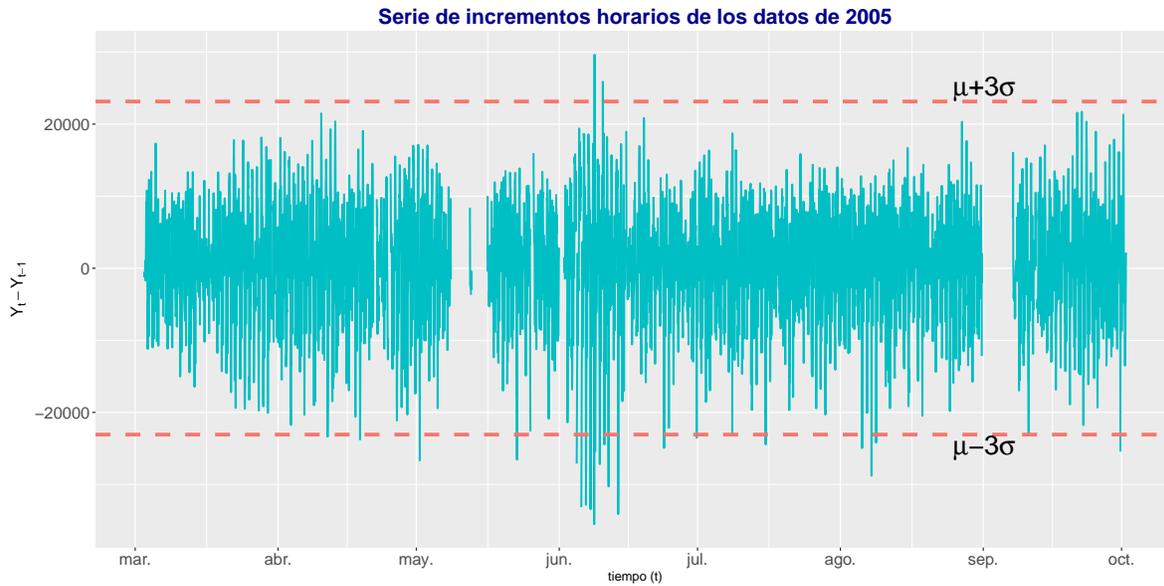


Figura 4.3: Incrementos de datos horarios en 2005 (marzo-septiembre).

La Figura 4.4 corresponde a los datos registrados durante 2009, se observa que febrero y diciembre son meses con menos del 70 % de datos para los tres primeros canales, mientras que el $S4_A$ tiene menos del 21 % de enero hasta agosto. Al analizar la figura, podemos notar que la distribución de los datos es bimodal, lo cual no es esperado; las distribuciones observadas se deben muy probablemente a que los datos que faltan son significativos. La figura 4.5 muestra que el canal $S3_A$ tiene variaciones mínimas de enero hasta julio, los otros canales presentan variaciones de 10 % hasta 20 % en ese intervalo de tiempo. A partir de agosto, se observa un aumento gradual en las cuentas, con variaciones de 8 % y 10 %, aproximadamente. De acuerdo a la tabla 4.2, en los dos primeros canales la distribución es asimétrica, sesgada a la derecha. Mientras que el $S1_A$ y $S4_A$ son los que presentan menor dispersión con respecto a la media.

	Mínimo	1er cuartil	Mediana	Media	3er cuartil	Máximo	desv.est
S1_A	0.5184	0.9232	1.000	1	1.0439	2.0360	0.07
S2_A	0.5498	0.8250	1.0698	1	1.0971	2.2183	0.14
S3_A	0.5678	0.9074	0.9148	1	1.1275	2.2919	0.12
S4_A	.481	0.952	0.986	1	1.031	1.986	0.06

Tabla 4.2: Resultados del análisis estadístico para los datos del año 2009.

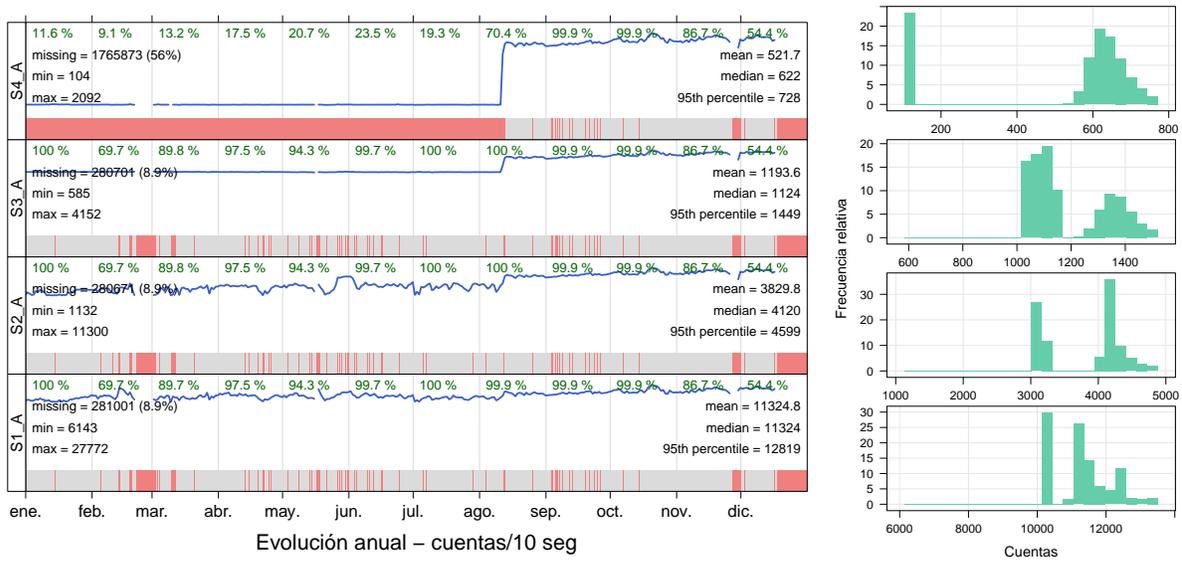


Figura 4.4: Comportamiento de datos para el año 2009.

Gráfica de datos diarios en 2009

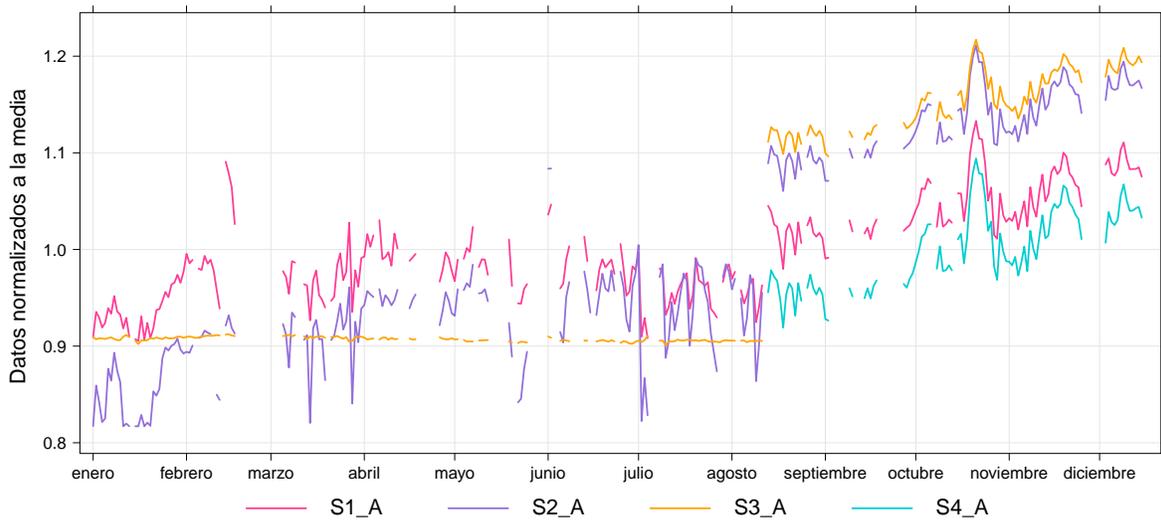


Figura 4.5: Datos normalizados a la media para 2009.

Posteriormente, para 2009 se realiza la prueba de control con el canal S1_A, el resultado se muestra en la figura 4.6. A pesar de la anomalía en las cuentas, producida en agosto para todos los canales, el canal S1_A tiene incrementos estables. Es probable que este incremento se deba a un error electrónico y que los datos después de agosto sean correctos bajo un factor de conversión.

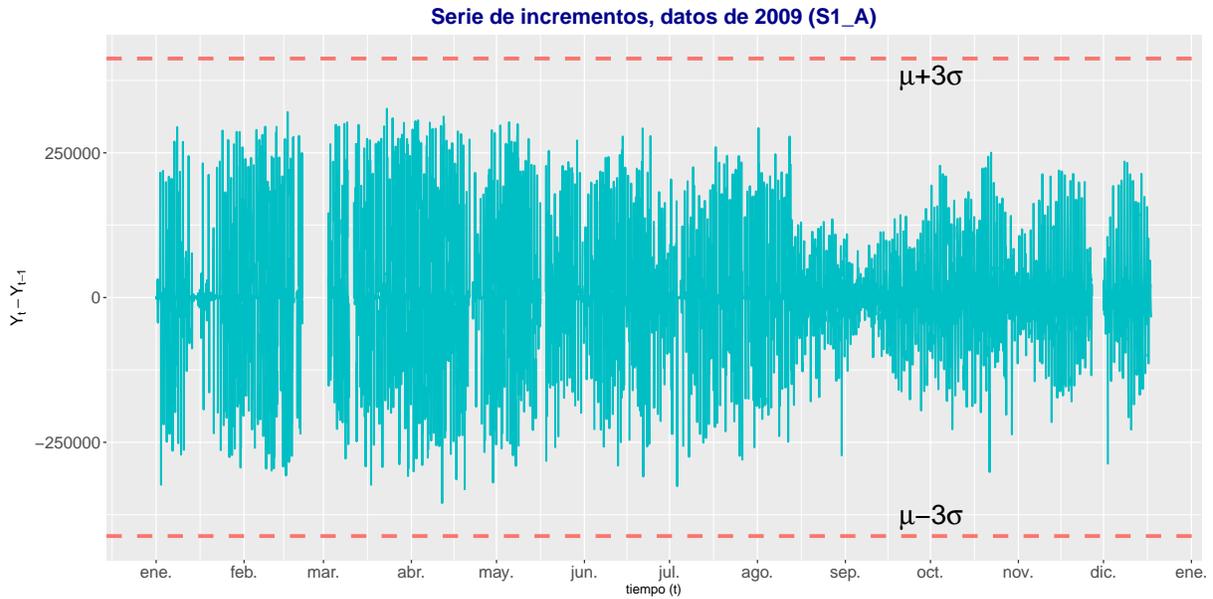


Figura 4.6: Incrementos de datos horarios en 2009.

Durante el año 2011, la figura 4.7 muestra que la distribución de los datos es casi simétrica debido a que hay días enteros sin registros y la desviación estándar está muy cerca de la media; sin embargo, los datos faltantes, la mayoría concentrados en febrero, abril, junio, noviembre y diciembre, suman 30 % del total para el año. En cuanto a la serie normalizada, figura 4.8, se observa que en las cuentas diarias hay una variación cercana al 15 % y una desviación estándar menor al 0.08. En la tabla 4.3, se puede corroborar que la distribución es casi simétrica con un pequeño sesgo a la derecha.

	Mínimo	1er cuartil	Mediana	Media	3er cuartil	Máximo	desv.est
S1_A	0.2680	0.9480	0.9891	1	1.0420	1.8949	0.07
S2_A	0.2716	0.9547	0.9898	1	1.0363	1.9612	0.06
S3_A	0.2760	0.9606	0.9916	1	1.0315	1.9884	0.05
S4_A	0.2747	0.9419	0.9867	1	1.0475	1.9226	0.08

Tabla 4.3: Resultados del análisis estadístico para los datos del año 2011.

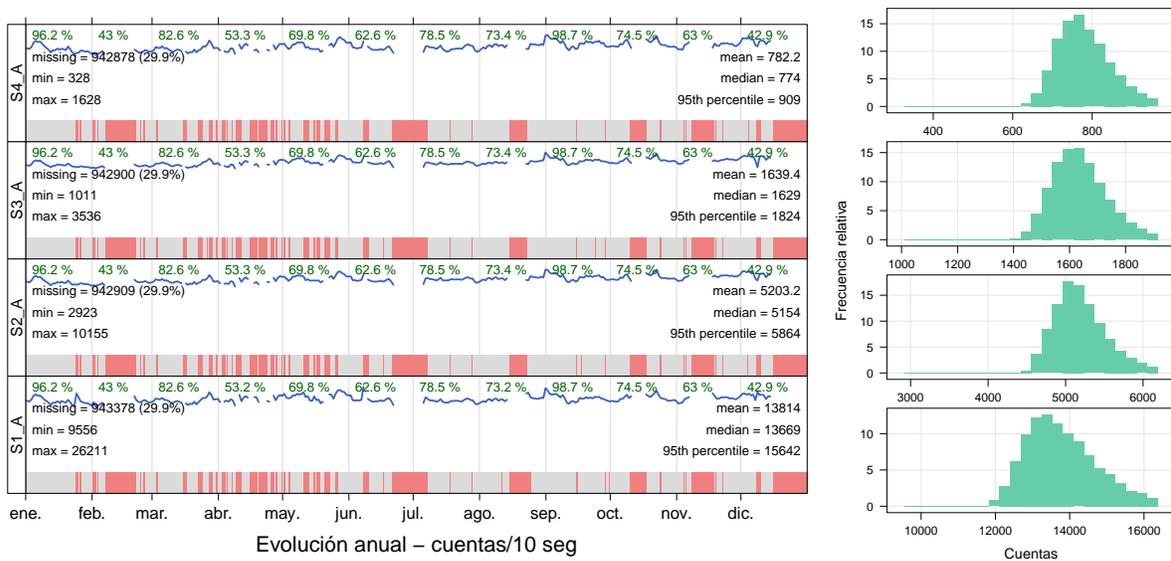


Figura 4.7: Comportamiento de datos para el año 2011.

Gráfica de datos diarios en 2011

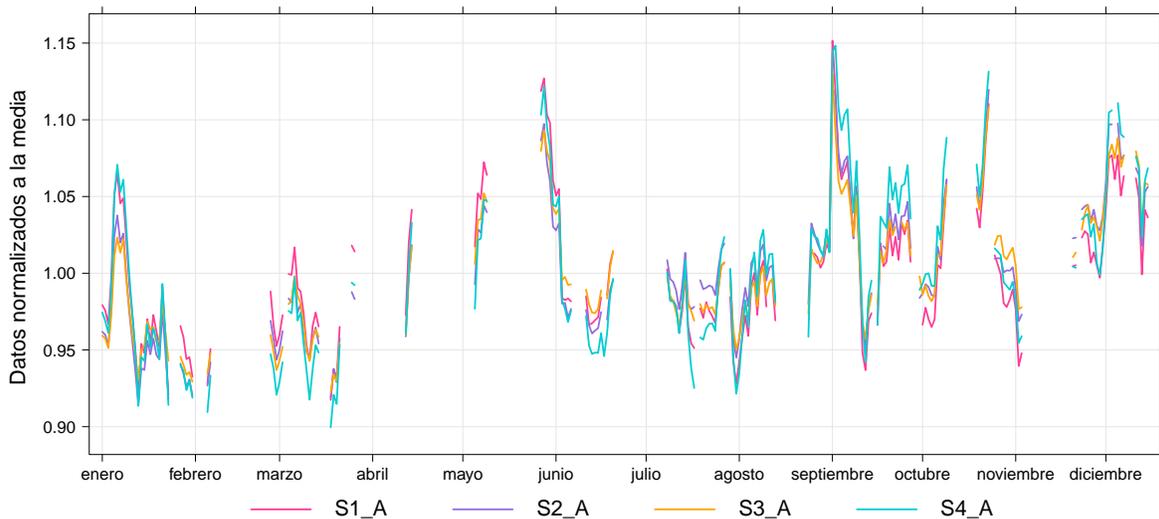


Figura 4.8: Datos normalizados a la media para 2011

A continuación se muestra el resultado de la prueba de estabilidad para los incrementos del canal $S1_A$ en la figura 4.9. Como se puede observar, sólo uno de los datos se sale de los límites de control, por lo tanto los incrementos del canal $S1_A$ son estables. Pero veamos cuál es la causa del pico entre agosto y septiembre en la figura 4.10, se observa que el 1 de septiembre la variación diaria no se comporta como debería, después esta variación va bajando sus cuentas, hasta recuperarse el 12 de septiembre. Sucede algo similar con el pico entre mayo y junio.

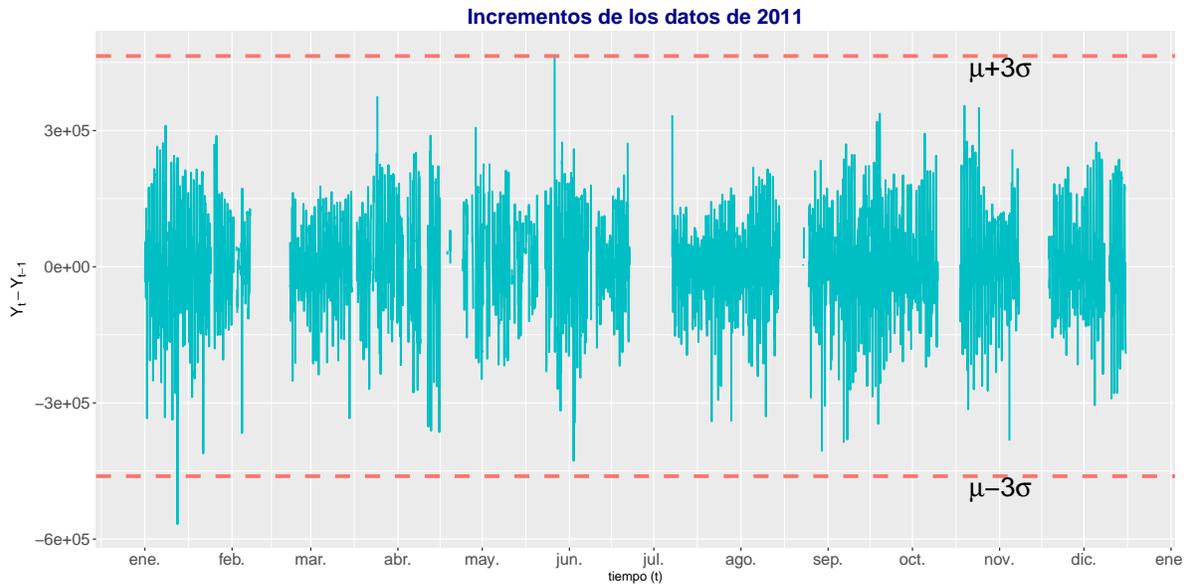


Figura 4.9: Incrementos de datos horarios en 2011.

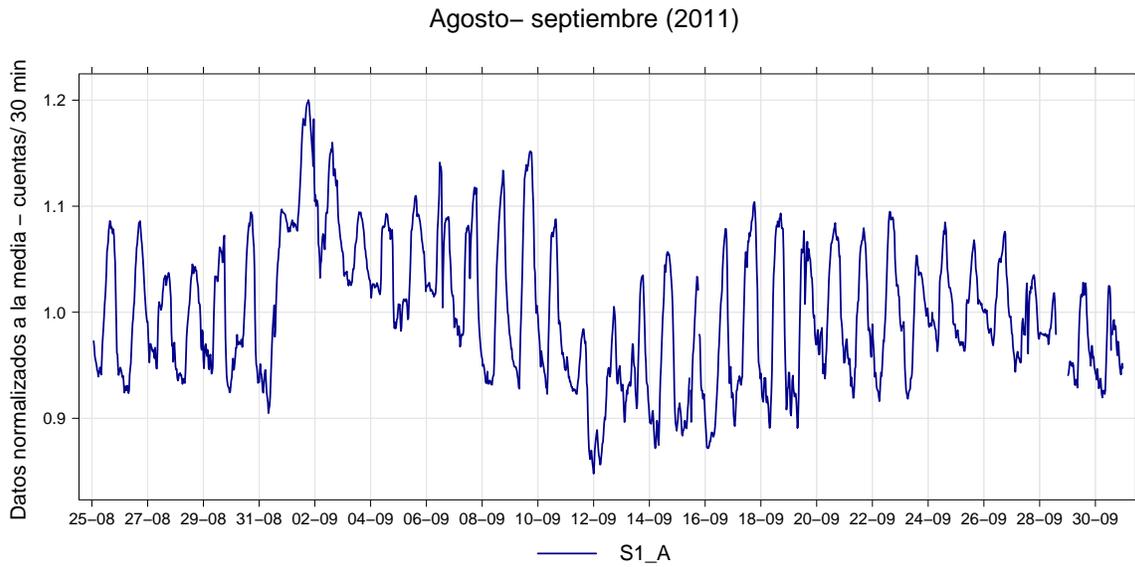


Figura 4.10: Variación diaria observada en agosto y septiembre de 2011.

Con base en la figura 4.11, para el año 2012 se tienen el 78% de los datos totales. Los meses de enero, julio, octubre y diciembre concentran la mayor cantidad de datos faltantes. Los histogramas para los tres primeros canales son casi simétricos con un ligero sesgo. En la serie normalizada de la figura 4.12, las variaciones que se observan son de aproximadamente 10% en los canales S1_A - S3_A y hasta 15% en el S4_A. Es importante resaltar que el comportamiento de los tres primeros canales son muy similares, mientras que el cuarto canal tiene datos más dispersos y su distribución está sesgada a la izquierda, como se muestra en la tabla 4.4.

	Mínimo	1er cuartil	Mediana	Media	3er cuartil	Máximo	desv.est
S1_A	0.1806	0.9572	0.9827	1	1.0292	1.9922	0.06
S2_A	0.1904	0.9625	0.9917	1	1.0270	2.0200	0.05
S3_A	0.1771	0.9667	0.9912	1	1.0225	2.0213	0.05
S4_A	0.2018	0.9821	1.0260	1	1.0530	2.1097	0.08

Tabla 4.4: Resultados del análisis estadístico para los datos del año 2012.

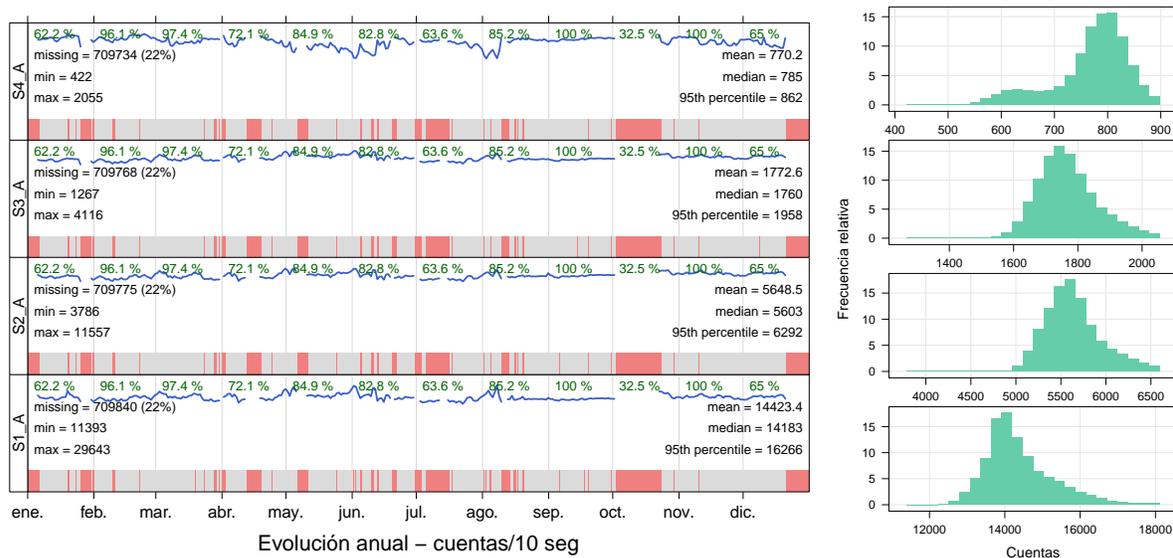


Figura 4.11: Comportamiento de datos para el año 2012.

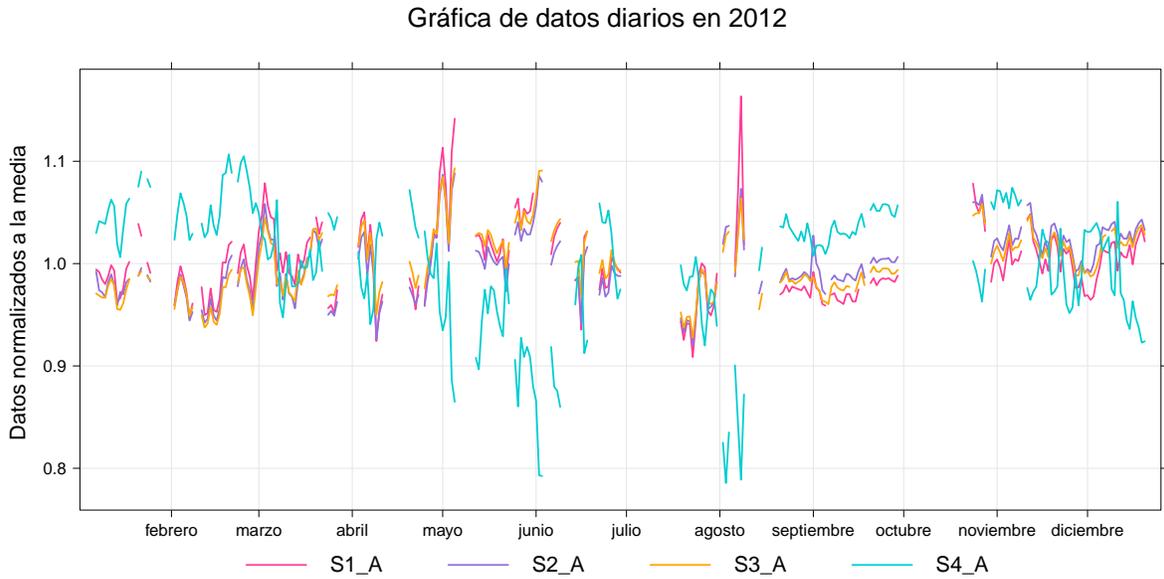


Figura 4.12: Datos normalizados a la media para 2012.

Como observamos en la gráfica anterior, el cuarto canal de energía tiene un alto grado de variabilidad, mientras que los tres primeros canales parecen tener cierta estabilidad y se comportan de forma similar. De esta manera, se grafica la serie de incrementos de datos horarios para el canal $S1_A$, véase la figura 4.13. Se muestra que los incrementos son estables, no hay aumentos repentinos; sin embargo, se observa que en mayo y agosto la diferencia en las cuentas es mayor y eso causa los picos en la gráfica normalizada.

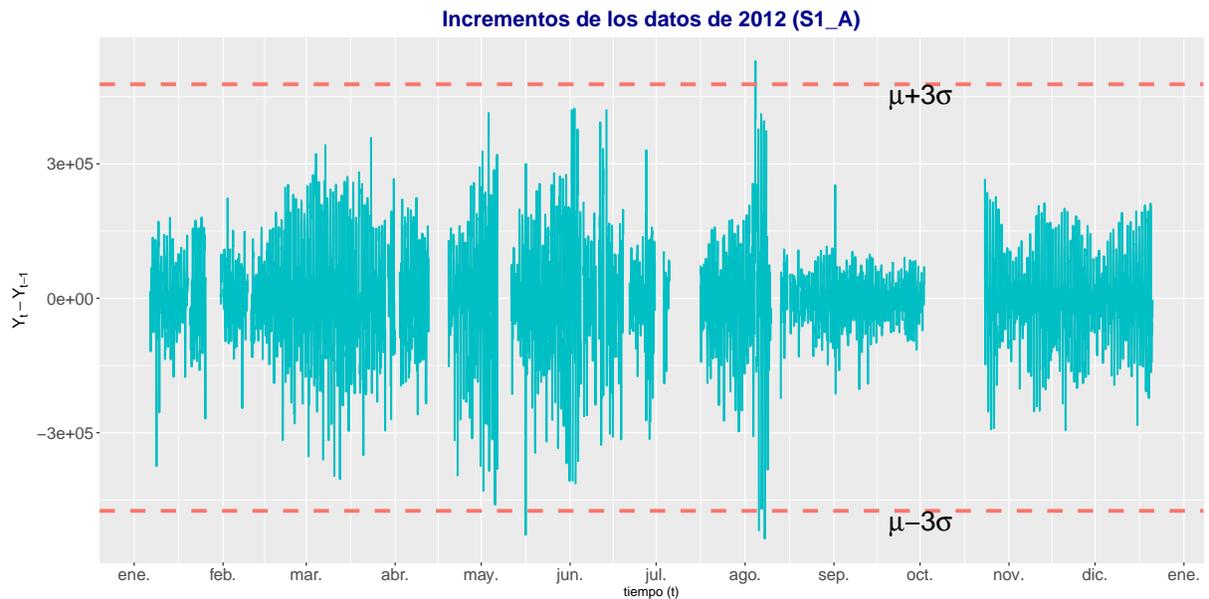


Figura 4.13: Incrementos de datos horarios en 2012.

En el año 2013 (ver figura 4.14), prácticamente no hay datos en enero, mientras que en septiembre, octubre y noviembre hay en promedio 53% de los datos totales. De este modo, para el año 2013 se tuvo un 28.5% de datos faltantes. Se puede ver también que en los histogramas hay sesgo a la izquierda tanto para los datos de 10 segundos como para los datos diarios (ver tabla 4.5) ya que la mediana es mayor que media. Sin embargo, no sucede lo mismo para el cuarto canal, el cual tiene una distribución bimodal que se debe probablemente a su gran variabilidad en el tiempo y errores en electrónica. En la serie normalizada de la figura 4.15 hay variaciones cercanas al 20% en el canal S4_A, mientras que el S1_A y S2_A presentan variaciones máximas del 10%. La señal del canal S3_A se muestra sólo para el mes de diciembre, porque para cada día hay al menos un dato faltante, por lo que el dato diario no es representativo.

	Mínimo	1er cuartil	Mediana	Media	3er cuartil	Máximo	desv.est
S1_A	0.0022	0.9699	0.9916	1	1.0310	1.4313	0.08
S2_A	0.0024	0.9828	1.0027	1	1.0290	1.1538	0.05
S3_A	0.002	0.998	1.018	1	1.041	1.142	0.07
S4_A	0.0027	0.8685	1.0053	1	1.1424	1.2499	0.14

Tabla 4.5: Resultados del análisis estadístico para los datos del año 2013.

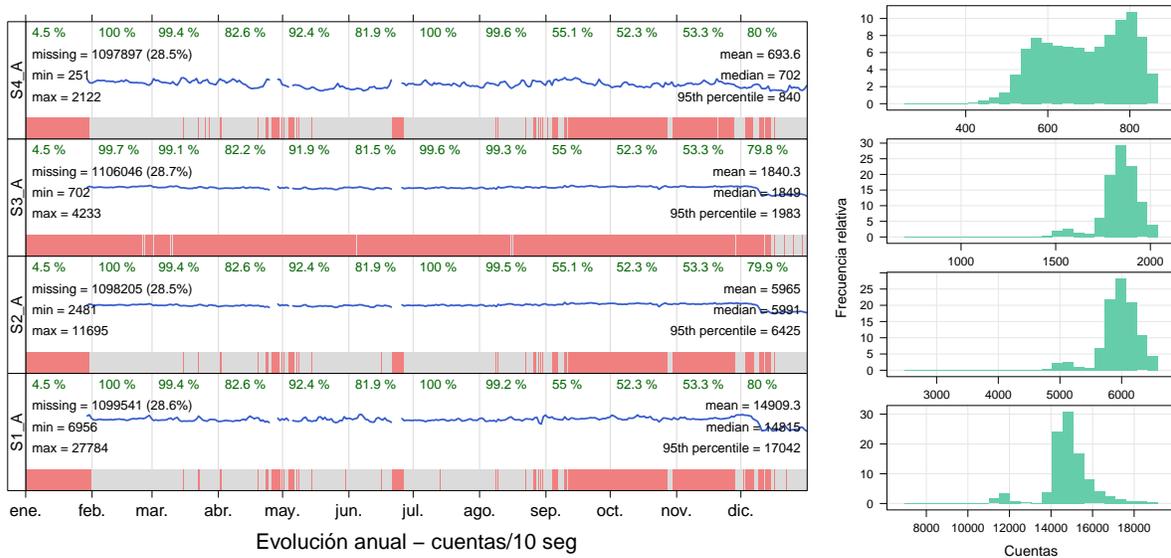


Figura 4.14: Comportamiento de datos para el año 2013.

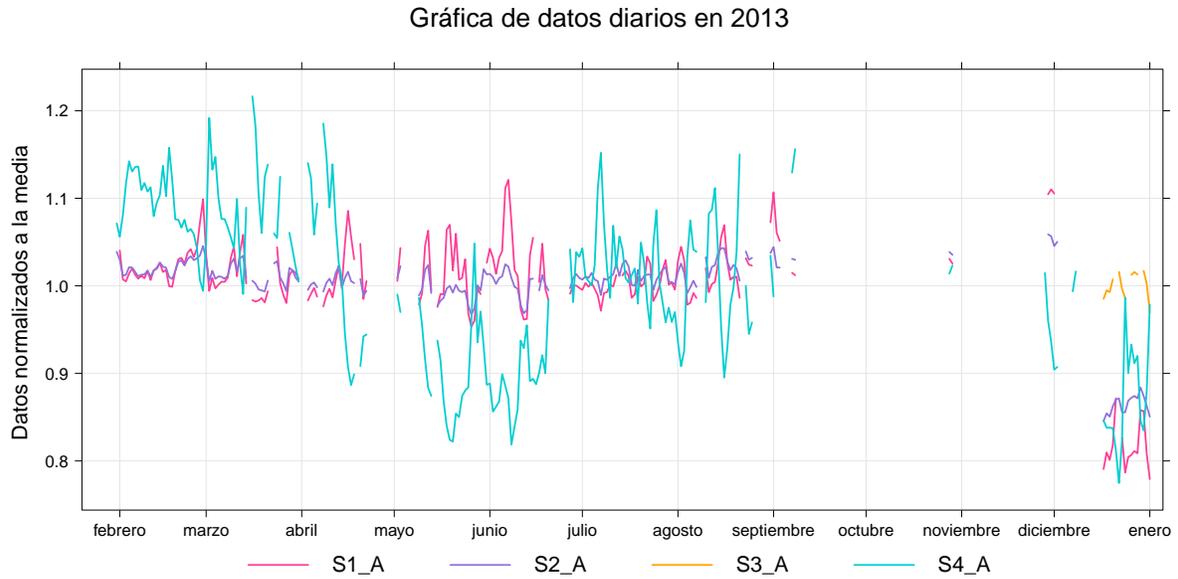


Figura 4.15: Datos normalizados a la media para el año 2013.

De igual forma, se grafica la serie de incrementos de los datos horarios (figura 4.16), esta vez para el canal $S2_A$, ya que el tercer canal está perdido, el cuarto canal ($S4_A$) muestra variaciones muy significativas y el $S1_A$ también muestra mucha variabilidad de abril a julio. Se observa que el porcentaje de datos que quedan fuera de los límites de control es menor al 5%, por lo tanto los incrementos en los datos horarios son estables.

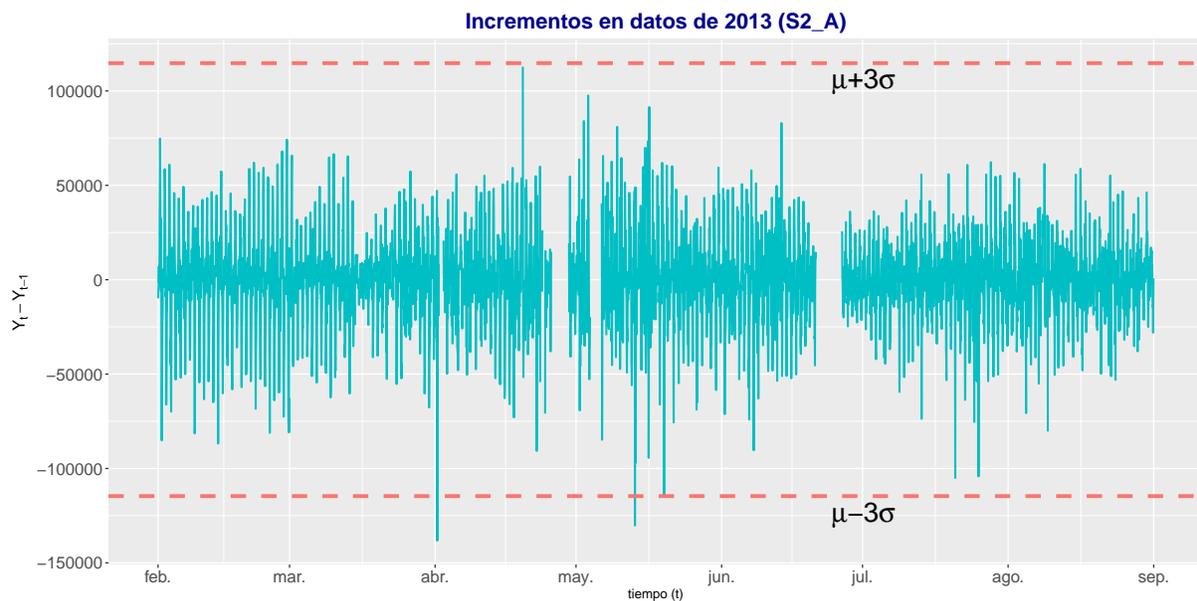


Figura 4.16: Incrementos de los datos horarios en 2013.

En el año 2014, la figura 4.17 muestra que se registró el 72 % de los datos, donde los meses de febrero, marzo y diciembre fueron los meses con menos registros (menor al 58 %). Debido a la falta del 28 % de los datos, la distribución es bimodal en los tres primeros canales; mientras que la distribución del $S4_A$ es debido a su alta variabilidad. Como se muestra en la figura 4.18, para la serie normalizada a la media, la variación es aproximadamente de 10 % para los tres primeros canales y cerca de 17 % para el $S4_A$, vemos que esto se corresponde con la desviación estándar que se muestra en la tabla 4.6. Nótese que en la serie diaria normalizada no hay datos en febrero y marzo, mientras que en la fig 4.18 sí, esto se debe a que en esos meses, todos los días tienen al menos un dato faltante.

	Mínimo	1er cuartil	Mediana	Media	3er cuartil	Máximo	desv.est
S1_A	0.003	0.953	0.977	1	1.045	1.397	0.13
S2_A	0.003	0.995	1.020	1	1.040	1.169	0.07
S3_A	0.003	0.998	1.018	1	1.037	1.104	0.07
S4_A	0.002	0.896	0.961	1	1.114	1.483	0.16

Tabla 4.6: Resultados del análisis estadístico para los datos del año 2014.

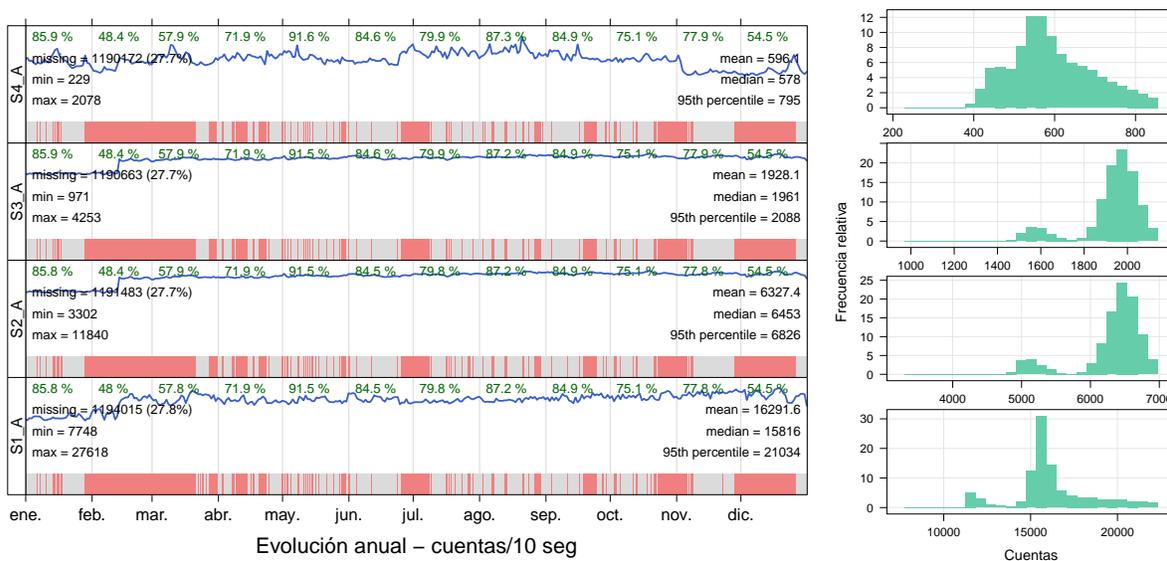


Figura 4.17: Comportamiento de datos para el año 2014.

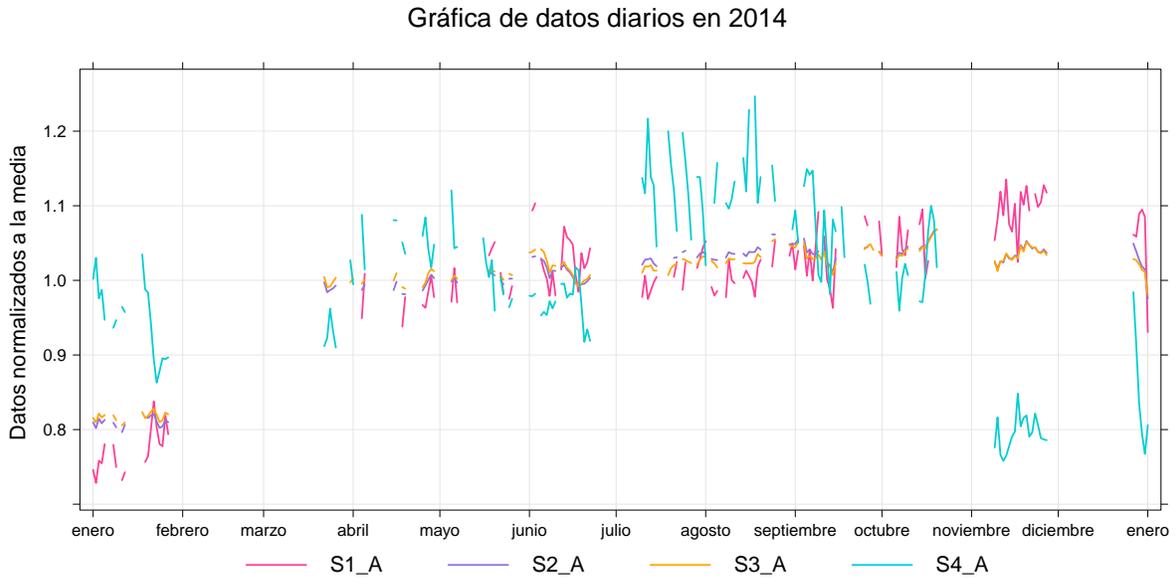


Figura 4.18: Datos normalizados a la media para el año 2014.

Posteriormente, se realiza la prueba de estabilidad para los incrementos de datos horarios del canal $S2_A$ de abril a noviembre (figura 4.19). Como no se observan incrementos repentinos en las cuentas en más del 5% de los datos, que superen los límites de control, el canal se considera que es estable.

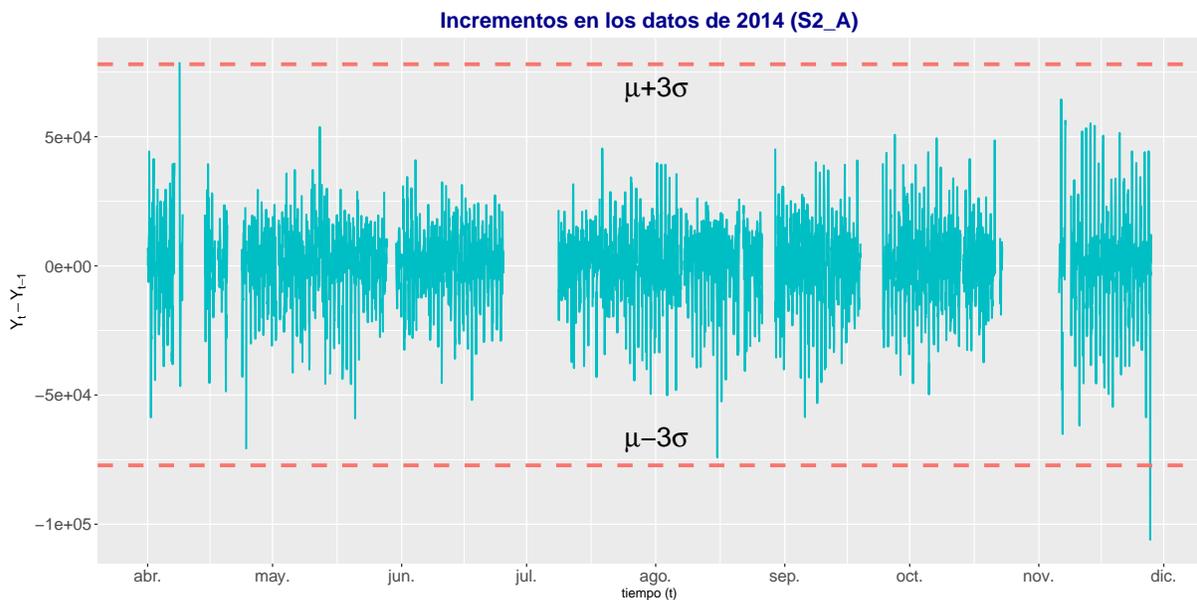


Figura 4.19: Incrementos de los datos horarios en 2014.

Las Figuras 4.20 y 4.21 muestran los resultados para el año 2015. De junio a octubre se registraron el menor porcentaje de datos. A finales de agosto se detuvo la toma de datos y se reinició en la segunda quincena de septiembre. Los canales $S3_A$ y $S4_A$ muestran distribuciones casi simétricas, mientras que el canal $S1_A$ y $S2_A$ son bimodales debido a su alta variabilidad. Se observa, también, que de febrero a junio, las variaciones del canal $S1_A$ son mayores que los registrados por los canales de mayor energía. Este comportamiento es consecuencia de los errores producidos en esas fechas y que se muestran en las figuras 4.22 y 4.23. Los resultados del análisis estadístico para los cuatro canales de deposición de energía se muestran en la tabla 4.7

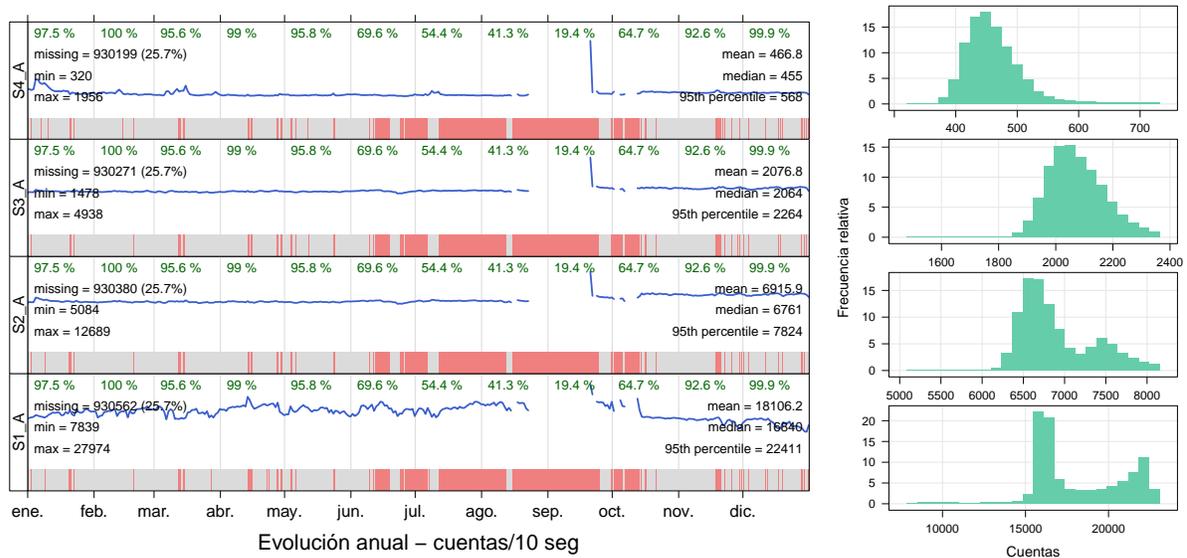


Figura 4.20: Comportamiento de datos para el año 2015.

	Mínimo	1er cuartil	Mediana	Media	3er cuartil	Máximo	desv.est
S1_A	0.0022	0.9003	0.9304	1	1.1364	1.8400	0.15
S2_A	0.0029	0.9438	0.9733	1	1.0636	2.1709	0.07
S3_A	0.0027	0.9625	0.9910	1	1.0311	2.0790	0.04
S4_A	0.0026	0.9258	0.9735	1	1.0289	2.0721	0.12

Tabla 4.7: Resultados del análisis estadístico para los datos del año 2015.

Los datos de los cuatro canales de energía están sesgados a la derecha y tienen dispersión con respecto a la media menor al 0.16. La falta de datos en la figura 4.21 que sí se muestran en la figura 4.20, para los meses de junio a noviembre se debe a que ninguno de esos días tiene el 100 % de observaciones.

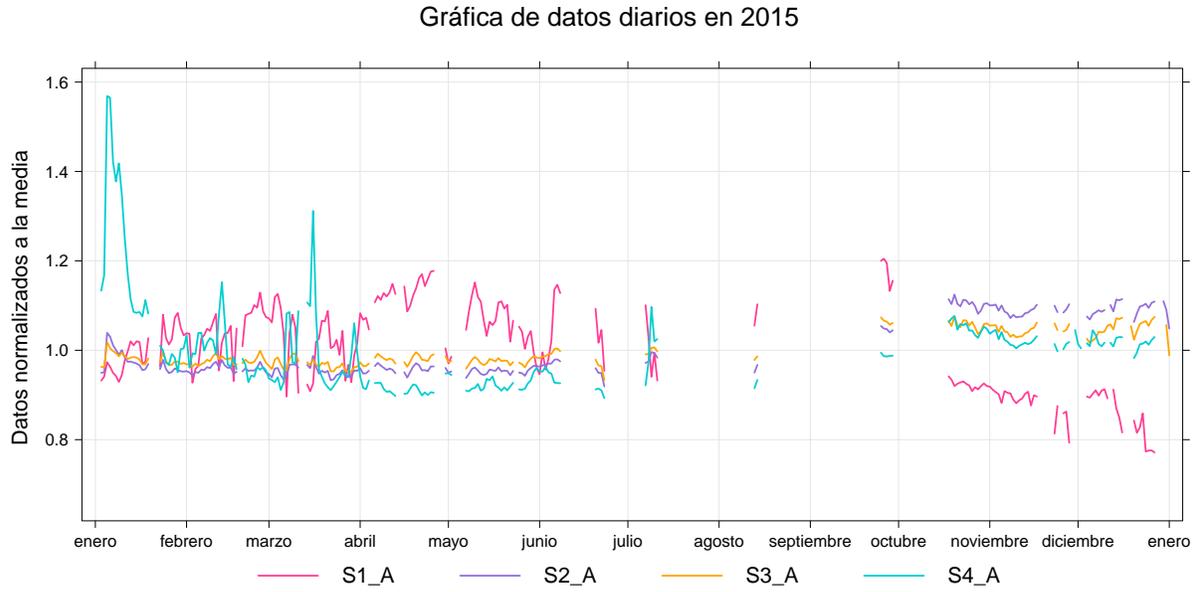


Figura 4.21: Datos normalizados a la media para el año 2015.

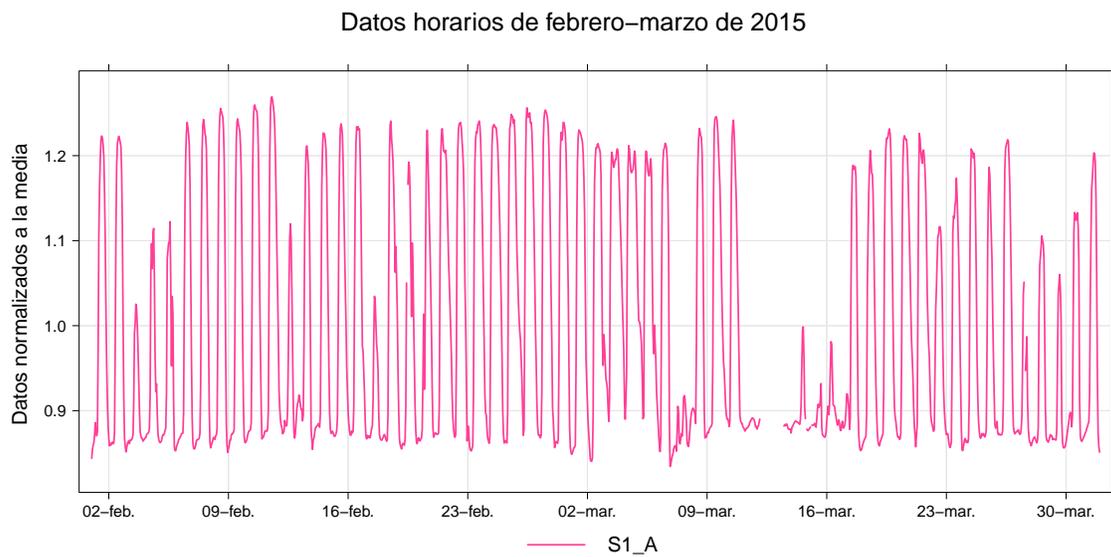


Figura 4.22: Datos horarios normalizados a la media en 2015 (Febrero - Marzo).

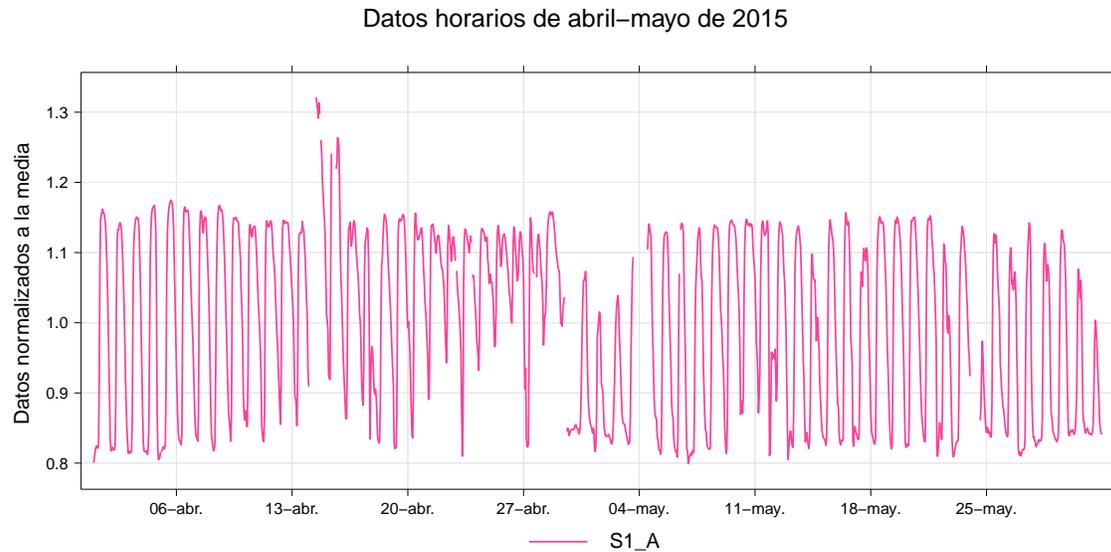


Figura 4.23: Datos horarios normalizados a la media en 2015 (Abril - Mayo).

Finalmente, se realiza la prueba de control estadístico para la serie diferenciada (incrementos) y se muestra en la figura 4.24 para el canal $S2_A$. En esta figura, se observa que de enero a junio, los datos son estables.

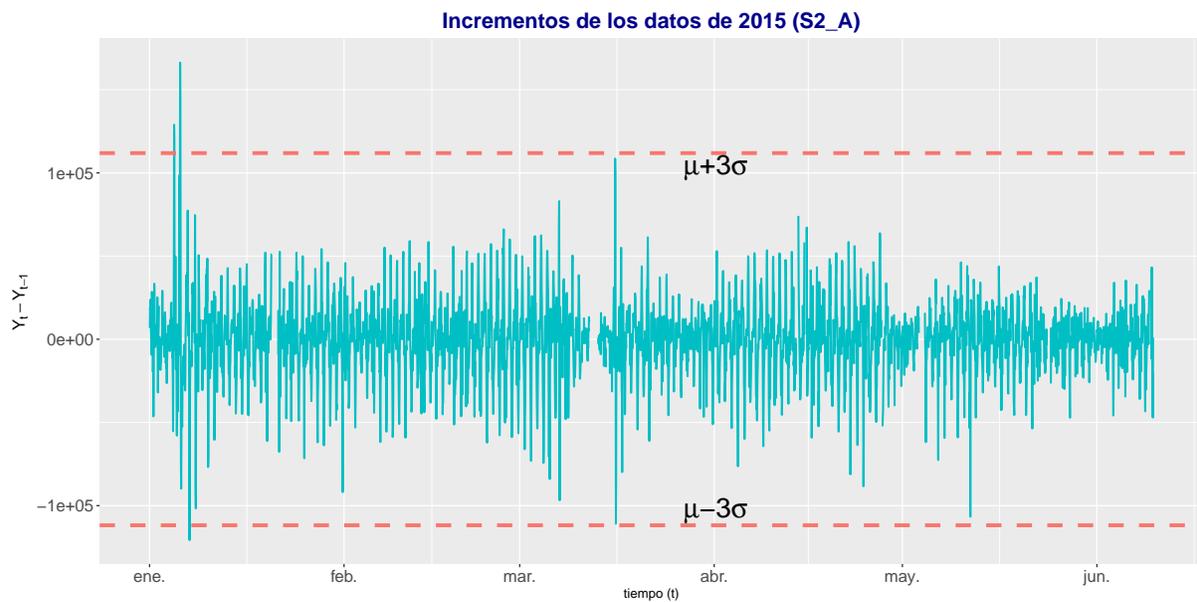


Figura 4.24: Serie diferenciada de datos horarios en 2015.

4.3. Señal total del TNS

Finalmente, en la figura 4.25 se muestra la señal total del detector normalizada a la media con razón de conteo de un día. En esta señal podemos observar, de forma general, cómo evoluciona el registro de los datos, los cortes y las variaciones. Notamos que en el año 2006 las cuentas disminuyeron, mientras que del 2009 al 2015 la serie tiene una tendencia creciente, presentando mayor variación en 2015. Esta variación no es lo esperado, es decir, se presenta un máximo de conteo durante el máximo de actividad solar y no al revés. Por lo tanto, deben existir otros factores que estén influyendo, como la electrónica o fuga de gas en los CP's.

Como se mencionó en el Capítulo 2, una de las variaciones importantes en el flujo de rayos cósmicos es la variación diurna. Para observar cómo se distribuye el máximo durante el día veamos la figura 4.26, donde se tomaron datos con razón de conteo 1 minuto y se normalizaron a la media. Esta gráfica nos muestra la hora en la que se encuentra el máximo de cada día durante los 13 años de detección, notamos que la mayoría se encuentra entre las 12 y 18 horas, lo cual concuerda de manera general con lo esperado para la variación diurna (ver figura 1.4).

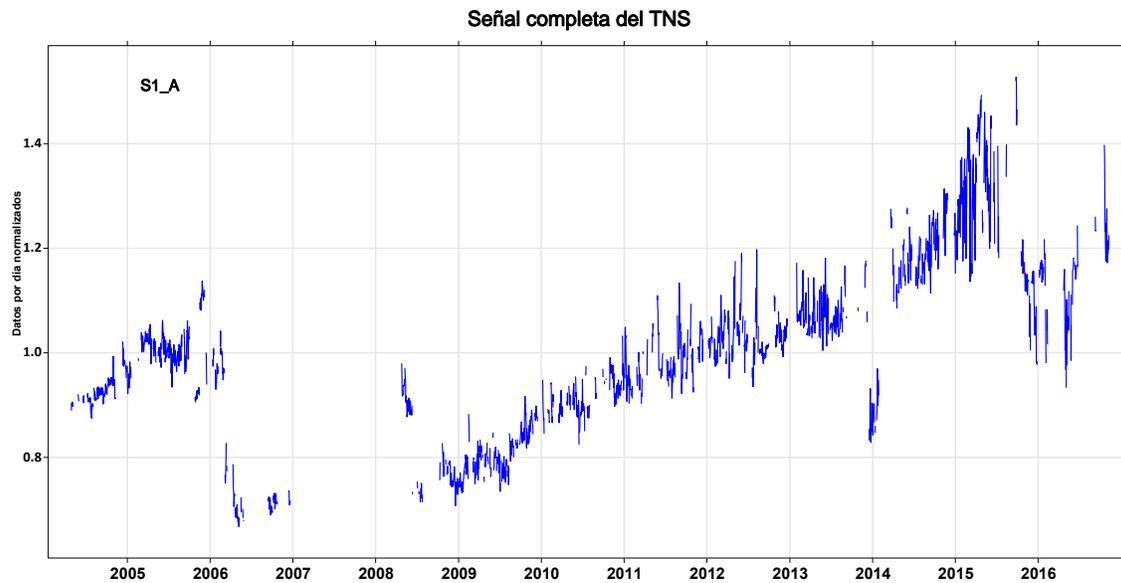


Figura 4.25: Señal completa para el canal *S1_A* con una razón de conteo de 24 horas.

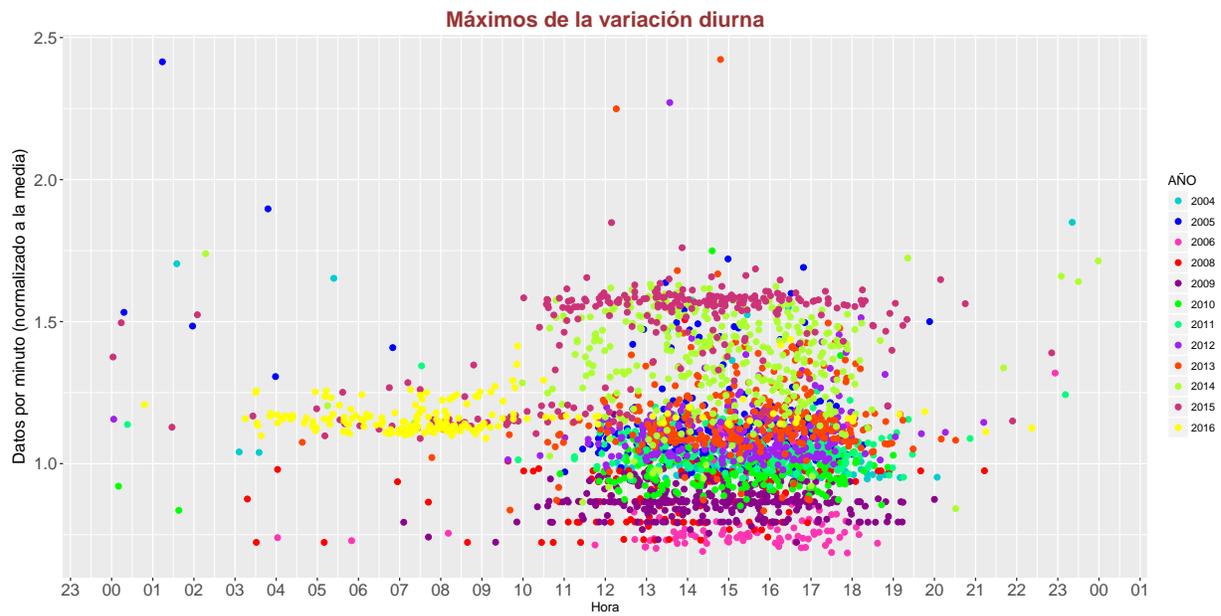


Figura 4.26: Máximos de la variación diaria, datos normalizados a la media. Se puede observar que la mayoría se encuentra entre las 12 y 18 horas, excepto para el año 2016 donde una parte se concentra entre las 6 y 9 horas.

4.4. Resultados

Con base en los análisis mostrados anteriormente, se puede concluir lo siguiente:

1. Los datos del año 2005 tienen un comportamiento estable de marzo a octubre en los cuatro canales de energía.
2. En 2009 los datos tienen incrementos estables; sin embargo, en la serie original hay un error electrónico que provoca un aumento en las cuentas a partir de agosto.
3. Para el año 2011, se observó que los incrementos son estables, aunque la serie original tiene varios errores electrónicos como se muestra en la figura 4.10. Estos errores provocaron que la serie muestre mucha variabilidad.
4. En el año 2012, sólo los tres primeros canales de energía son estables. El canal S1_A, muestra un porcentaje de error menor al 5%.
5. En 2013, el canal de energía S2_A se muestra estable.
6. Se observó que para el 2014 los canales S2_A y S3_A son estables de abril a diciembre.
7. Para la serie registrada en 2015, se muestra que los canales S2_A y S3_A son estables de enero a junio.
8. Las partículas neutras registradas por el TNS no muestran una anticorrelación con el ciclo de manchas solares.
9. El máximo de la variación diurna se encuentra, en su mayoría, entre las 12 y 18 horas, lo que es coherente con la variación diurna que registran otros detectores a nivel de tierra.

Conclusiones

En este trabajo, se realizó un análisis de estabilidad estadística de las partículas neutras registradas del 2004 al 2016, por el Telescopio de Neutrones 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 ayuda a conocer los datos a descartar y en qué intervalo de tiempo se requiere hacer un análisis más detallado si se requiere extraer información física de los mismos; además, facilita el análisis para estudios de física solar. 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, por lo que 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.

También conocemos las variaciones por canal de energía y podemos detectar qué canal tiene problemas de software, hardware o de sistema de adquisición de datos. El cuarto canal de deposición de energía (*S4_A*), ha presentado variaciones más altas con respecto a los otros tres canales (*S1_A*, *S2_A* y *S3_A*). Por otro lado, el canal *S3_A* es el que ha tenido más errores electrónicos, por ejemplo en el año 2013, la mayoría de los registros son datos constantes, que no reflejan influencia de fenómenos físicos. Por lo tanto, en todo este tiempo de observación del TNS, el segundo canal de energía es el más estable. Con base en el análisis de estabilidad realizado, se conoce el porcentaje de datos faltantes para cada mes y estadísticos para la resolución mínima de conteo; además de conocer el comportamiento general de los datos diarios registrados. Los errores detectados 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.

Se obtuvo la señal total del detector. Esto nos muestra cómo se ha comportado el TNS a lo largo del tiempo de análisis. Con base en el continuo registro de datos por parte del TNS para los años siguientes, se analizará si la tendencia ascendente se mantiene o bajará eventualmente. Esta tendencia en los registros, sirve como precursor para un análisis de los diferentes componentes del TNS; de este modo, se puede medir la presión del gas contenido en los contadores proporcionales y conocer si existe fuga de gas, que pueda influenciar en los registros de datos. También se puede analizar el sistema de adquisición de datos y tarjetas electrónicas, para descartar y, en su caso, arreglar posibles desperfectos que generen variaciones en los registros del

TNS.

Con base en la figura 4.26 se pudo observar que el máximo de la variación diurna se encuentra, en su mayoría, entre las 12 y 18 horas, lo que es consistente con la variación diurna en los observatorios a nivel de tierra.

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.

De este modo, podemos concluir que el detector funciona de forma adecuada en términos generales. Ha detectado dos eventos de neutrones solares y tiene un registro del 86 % de datos desde su puesta en marcha en junio de 2004.

Imágenes de Actividad Solar

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 a 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 [34]. 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, se recomienda ver las fotos y animaciones en la pestaña de galerías que se encuentra accediendo al enlace de cada imagen.

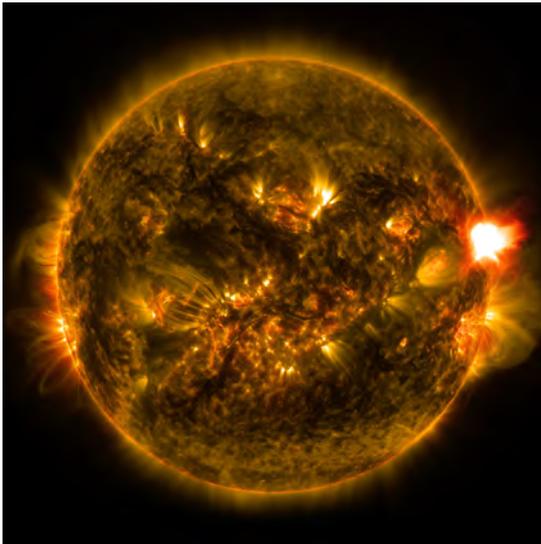


Figura A.1: El Sol emitió una fulguración solar de clase M, alcanzando su punto máximo a las 11:24 p.m. el 12 de enero de 2015. El Observatorio de Dinámica Solar de la NASA, que observa el sol constantemente, capturó una imagen del evento^a.

¹<https://www.nasa.gov/content/goddard/nasa-releases-images-of-1st-notable-solar-flare-of-2015>

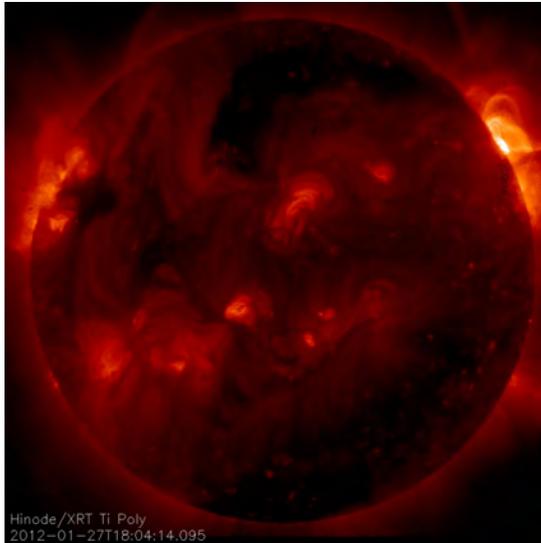


Figura A.2: El 27 de enero de 2012, estalló una fulguración intensa de clase X. Las fulguraciones de clase X son las explosiones más intensas. Aquí se ve una imagen de la llamarada capturada por el telescopio de rayos X en Hinode. Esta imagen muestra una emisión de plasma calentado a más de ocho millones de grados durante el proceso de liberación de energía de la llamarada^a.

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 fotosfera solar, tomada por LACIGE, ENES-UNAM unidad Morelia.



²Imagen tomada de https://www.nasa.gov/multimedia/imagegallery/image_feature_2169.html

Referencias

- [1] María Guadalupe Cordero Tercero, Alejandro Lara Sánchez, María Dolores Maravilla Meza, Blanca Emma Mendoza Ortega, y José Francisco Valdés Galicia. *Introducción a la física espacial*. Universidad Nacional Autónoma de México, 2013. VII, 3, 6
- [2] Javier A. Otaola y José Fco. Valdés-Galicia. *Los rayos cósmicos: mensajeros de las estrellas*. Fondo de Cultura Económica, 1992. IX, XIII, 1, 2, 7
- [3] Manuel Aguilar Benitez de Lugo. Cien años de rayos cósmicos. *Rev. R. Acad. Cienc. Exact. Fís. Nat.*, 2011. XIII, 1
- [4] Bruno B. Rossi. *Cosmic Rays*. McGraw-Hill, 1964. XIII
- [5] Hansjörg Schlaepfer. Cosmic Rays. *The Century of Space Science.*, 2003. XIII
- [6] Luis Xavier González. *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. XIII, XIV, 7, 11
- [7] S Shibata. Propagation of solar neutrons through the atmosphere of the earth. *Journal of Geophysical Research: Space Physics*, 99(A4):6651–6665, 1994. XIII
- [8] K. Watanabe, M. Gros, P. H. Stoker, K. Kudela, C. Lopate, J. F. Valdés-Galicia, A. Hurtado, O. Musalem, R. Ogasawara, Y. Mizumoto, M. Nakagiri, A. Miyashita, Y. Matsubara, T. Sako, Y. Muraki, T. Sakai, y S. Shibata. Solar Neutron Events of 2003 October–November. *The Astrophysical Journal*, 636(2):1135, 2006. XIV
- [9] X-M Hua y RE Lingenfelter. Solar flare neutron production and the angular dependence of the capture gamma-ray emission. *Solar Physics*, 107(2):351–383, 1987. XIV
- [10] Instituto de Geofísica, UNAM. www.geofisica.unam.mx/observatorios/rayos_cosmicos/grupo_raycos/antecedentes.html, 09 de Marzo de 2017. XIV, XIV, 9
- [11] JF Valdes-Galicia, A. Hurtado, O. Musalem, LX Gonzalez, Y. Muraki, Y. Matsubara, T. Sako, K. Watanabe, N. Hirano, N. Tateiwa, H. Tsujihara, Shoichi Shibata, y Taku Sakai. A New Solar Neutron Telescope in Mexico. In *International Cosmic Ray Conference*, volume 6, page 3433, 2003. XIV, 7

- [12] LX Gonzalez, JF Valdés-Galicia, A Hurtado, O Musalem, Y Muraki, T. Sako, K. Watanabe, Y. Matsubara, S. Shibata, T. Sakai, y H. Tsujihara. Energy spectrum for the solar neutron event of september 7 2005, derived from the SNT at Sierra Negra. In *Proceedings of the 30th international cosmic ray conference*, volume 1, pages 57–60, 2008. [xiv](#)
- [13] Y. Muraki, D. Lopez, K. Koga, F. Kakimoto, T. Goka, L. X. González, S. Masuda, Y. Matsubara, H. Matsumoto, P. Miranda, O. Okudaira, T. Obara, J. Salinas, T. Sako, S. Shibata, R. Ticona, Y. Tsunesada, J. F. Valdés-Galicia, K. Watanabe, y T. Yamamoto. Simultaneous Observation of Solar Neutrons from the International Space Station and High Mountain Observatories in Association with a Flare on July 8, 2014. *Solar Physics*, 291(4):1241–1265, Apr 2016. [xiv](#)
- [14] Thomas K. Gaisser. *Cosmic Rays and Particle Physics*. Cambridge University, first edition edition, 1990. [1](#)
- [15] LX González, JF Valdés-Galicia, F. Sánchez, Y. Muraki, T. Sako, K. Watanabe, Y. Matsubara, Y. Nagai, S. Shibata, T. Sakai, O. Musalem, y A. Hurtado. Re-evaluation of the neutron emission from the solar flare of 2005 september 7, detected by the solar neutron telescope at sierra negra. *The Astrophysical Journal*, 814(2):136, 2015. [1](#)
- [16] Hansjorg Schlaepfer. Cosmic rays. *Spatium*, 11:3–15, 2003. [2](#)
- [17] Peter KF Grieder. *Cosmic rays at Earth*. Gulf Professional Publishing, 2001. [4](#), [6](#)
- [18] Todor Stanev. *High energy cosmic rays*. Springer Science & Business Media, 2010. [4](#)
- [19] JF Valdés-Galicia, Y. Muraki, K. Watanabe, Y. Matsubara, T. Sako, LX Gonzalez, O. Musalem, y A. Hurtado. Solar neutron events as a tool to study particle acceleration at the sun. *Advances in Space Research*, 43(4):565–572, 2009. [7](#)
- [20] Y. Muraki, Y. Matsubara, I. Imaida, T. Koi, T. Murata, A. Okada, S. Sakakibara, T. Sako, H. Tsuchiya, Y. Munakata, et al. Solar flare and neutron telescope. *Nuclear Physics B-Proceedings Supplements*, 60(3):3–11, 1998. [7](#)
- [21] H. Tsuchiya, Y Muraki, K. Masuda, Y. Matsubara, T. Koi, T. Sako, S. Ohno, T. Hoshida, S. Shibata, Y. Munakata, 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):183–193, 2001. [7](#), [8](#)
- [22] Blanca Mendoza-Ortega et al. Chismes espaciales, 2001. [8](#)
- [23] Ross-Ihaka y Robert-Gentleman. R: a language for data analysis and graphics. *Journal of Computational and Graphical Statistics* 5: 299–314, 1996. [17](#)
- [24] Jorge A Ahumada. R para principiantes. *University of Hawaii*, 2003. [17](#), [18](#)
- [25] Julio Sergio Santana y Efrain Mateos. *El arte de programar en R: un lenguaje para la estadística*. México: Instituto Mexicano de Tecnología del Agua. UNESCO. Comité Nacional Mexicano del Programa Hidrológico Internacional., 2014. [19](#)

-
- [26] Hadley Wickham. Tidy data. *The Journal of Statistical Software*, 59, 2014. 21
- [27] Hadley Wickham, Romain Francois, Lionel Henry, y Kirill Müller. *dplyr: A Grammar of Data Manipulation*, 2017. R package version 0.7.1. 25
- [28] Garrett Grolemund y Hadley Wickham. Dates and times made easy with lubridate. *Journal of Statistical Software*, 40(3):1–25, 2011. 25
- [29] David C. Carslaw y Karl Ropkins. openair — an r package for air quality data analysis. *Environmental Modelling and Software*, 27–28(0):52–61, 2012. 27
- [30] Fernando Follos Pliego. *Lenguaje R aplicado al análisis de Calidad de Aire*. Environmental Risk Management, Septiembre de 2012. 27
- [31] Robert H. Shumway y David S. Stoffer. *Time Series Analysis and Its Applications*. Springer, 3 edition. 29
- [32] William W.S. Wei. *Time Series Analysis: Univariate and Multivariate Methods*. Pearson Education, segunda edición edition. 30
- [33] Richard A. Johnson y Gouri K. Bhattacharyya. *Statistics, Principles & Methods*. John Wiley & Sons, 6 edition. 30
- [34] NASA. https://www.nasa.gov/mission_pages/sunearth/news/X-class-flares.html, 20 de Marzo de 2017. 53
- [35] David Kahle y Hadley Wickham. ggmap: Spatial visualization with ggplot2. *The R Journal*, 5(1):144–161, 2013.