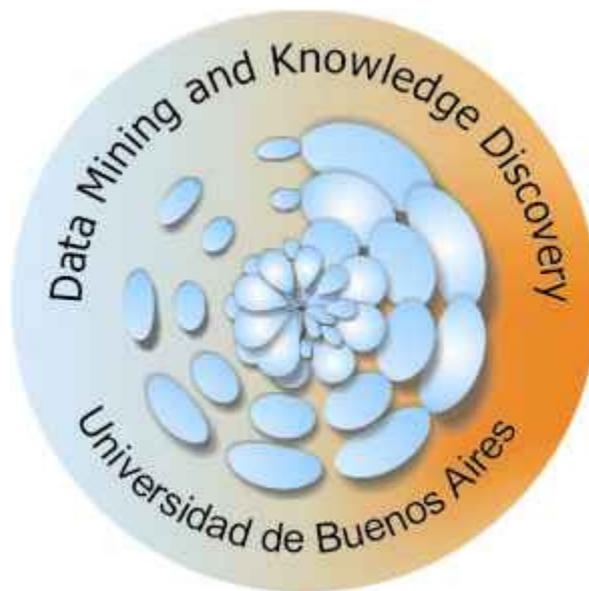


Universidad de Buenos Aires - FCEN y FiUBA

Maestría en Explotación de Datos

y

Descubrimiento del Conocimiento



Trabajo Final de Especialización

**Análisis sobre la ponderación de satélites en
comparaciones de Vista Común GPS**

Año: 2017

Alumno: Leandro Martín Ríos

Supervisión: Dr. Marcelo Soria

Índice

Resumen	4
Introducción.....	5
Material y métodos	7
Vista común cada 144 minutos.....	7
Vista común cada 16 minutos	9
Vista común	10
Análisis detallado de Vista Común	13
Limpieza de datos.....	15
Limpieza de datos - Sistema de Medición 1	15
Limpieza de datos - Sistema de Medición 2	20
MAD una medida de dispersión robusta	25
Generando un promedio ponderado.....	27
Comprobación de los supuestos del modelo	32
Un modelo por cada sistema de medición	37
Sistema de medición 1	37
Comprobación de los supuestos Sistema de Medición Nro 1.....	37
Sistema de Medición Nro. 2	39
Comprobacion de los supuestos Sistema de medición Nro. 2.....	39
Aplicación de ambos modelos.....	41
Resultados.....	43
Conclusiones	44
Agradecimientos	44
Referencias	44

Índice de tablas

Tabla 1. Datos de VC 144 minutos.....	7
Tabla 2. Datos crudos del Sistema de Medición 1	10
Tabla 3. Detalle de las variables del conjunto de datos	11
Tabla 4. Estadísticos descriptivos VC 16 min.....	12
Tabla 5. Estadísticos descriptivos VC 144 min.....	12
Tabla 6. Estadísticos descriptivos de variables de interés del Sist. de Medición Nro1	18
Tabla 7. Detalle de los datos con MDTR = 45,5 ns.....	19
Tabla 8. Estadísticos descriptivos de variables de interés del Sist. de Medición Nro. 2	22
Tabla 9. Tabla con resultados de los test de normalidad	25
Tabla 10. Variables de interés para la Reg. Lineal Múltiple	27
Tabla 11. Comparación de promedios en VC 16 min.....	28
Tabla 12. Primero modelo de Regresión Lineal Múltiple.....	28
Tabla 13. Segundo modelo de Regresión Lineal Múltiple.....	29
Tabla 14. Modelo de Regresión, con interacciones tomadas de a dos.....	30
Tabla 15. Modelo de Doble Interacción.	31
Tabla 16. Modelo de Regresión mediante la técnica "stepwise"	31

Tabla 17. Betas estandarizados del Modelo Doble Interacción.....	36
Tabla 18. Modelo de Regresión para el del Sist. de Med. Nro. 1.....	37
Tabla 19. Estadísticos descriptivos de los residuos estudentizados del Modelo del Sist. de Med. 1.....	37
Tabla 20. Modelo de Regresión para el del Sist. de Med. Nro. 2.....	39

Índice de figuras

Figura 1. Esquema de Medición Directa	5
Figura 2. Esquema de Vista Común	5
Figura 3. Diferencias de fase entre INT2 e INTI1 en VC 144 min durante seis días.....	7
Figura 4. Comparación ente Vista Común cada 144 minutos y Medición Directa	8
Figura 5. A izq.: VC 144 min y Med. Directa -superpuestas-. A derecha: Desvío de Allan	8
Figura 6. Evolución en el tiempo de los diferencias de fase entre INT2 e INTI1 en VC 16 minutos	9
Figura 7. Desvío de Allan para Vista Común cada 16 minutos.	10
Figura 8. Comparación de VC 144min y VC 16 min.	12
Figura 9. Día 16 de Marzo 2017 en VC 144 min.....	13
Figura 10. Día 16 de Marzo 2017 en VC 144 min. y sus medidas de VC 16 min.....	13
Figura 11. Detalle 2017-03-16 en VC 144 min, sus medidas de VC 16 min con incertidumbres asociadas.	14
Figura 12. Med. Nro 1 del 2017-03-16. Valores promedio de VC 16 min. y Vista Común.....	15
Figura 13. Diagrama de caja para el Sistema de Medición 1	16
Figura 14. Diagrama de caja para el Sistema de Medición 1 - variables normalizadas -	16
Figura 15. Gráfico de densidad de kernel para el Sistema de Medición Nro. 1	17
Figura 16. Extensivo correlograma para el Sistema de Medición Nro. 1	17
Figura 17. Correlograma para el Sistema de Medición Nro. 1	18
Figura 18. ELV_Reloj1 vs. DSG_Reloj1	18
Figura 19. ELV_Reloj1 vs. MDTR_Reloj1	19
Figura 20. MDTR_Reloj1 vs. DSG_Reloj1.....	19
Figura 21. MDIO_Reloj2 vs. DSG_Reloj1.....	20
Figura 22. Diagrama de caja para el Sistema de Medición Nro. 2	20
Figura 23. Diagrama de caja para el Sistema de Medición 2 - variables normalizadas -	21
Figura 24. Gráfico de densidad de kernel para el Sistema de Medición Nro. 2	21
Figura 25. Correlograma para el Sistema de Medición Nro. 2	22
Figura 26. ELV_Reloj2 vs. DSG_Reloj2	23
Figura 27. ELV_Reloj2 vs. MDTR_Reloj2.....	23
Figura 28. MDTR_Reloj2 vs. DSG_Reloj2.....	23
Figura 29. MDIO_Reloj2 vs. DSG_Reloj2.....	24
Figura 30. MSIO_Reloj2 vs DSG_Reloj2.....	24
Figura 31. Diagrama de caja de ISG_Reloj2	24
Figura 32. Densidad de Kernel de ISG_Reloj2	24
Figura 33. DSG_Reloj2 vs ISG_Reloj2.....	24
Figura 34. Densidad de Kernel, antes y después de la limpieza- 6 días-	26
Figura 35. Densidad de Kernel, antes y después de la limpieza- 16 marzo 2017-.....	26
Figura 36. Regresión lineal simple.....	27
Figura 37. Valores ajustados vs. Residuos estudentizados del Modelo Doble Interacción.....	32
Figura 38. Variables predictoras vs. Residuos estudentizados del Modelo Doble Interacción.....	33
Figura 39. Histograma de los residuos estudentizados del Modelo Doble Interacción	33
Figura 40. Gráfico Q-Q de Histograma de los residuos estudentizados del Modelo Doble Interacción	33
Figura 41. Distancias de Cook para el Modelo Doble Interacción.....	34
Figura 42. Correlograma de las variables del Modelo Doble Interacción.....	35
Figura 43. Valores de VIF variables del Modelo Doble Interacción.....	36
Figura 44. PCA sobre variables del Modelo Doble Interacción	36
Figura 45. Residuos estudentizados vs. Valores ajustados para Modelo del Sistema de Medición 1	37
Figura 46. Diagrama de caja de los residuos estudentizados para Modelo del Sistema de Medición 1.....	37

Figura 47. Histograma de los residuos estudentizados del Modelo del Sistema de Medición Nro. 1	38
Figura 48. Gráfico Q-Q residuos estudentizados del Modelo del Sistema de Medición Nro. 1.....	38
Figura 49. Correlograma de las variables Modelo del Sistema de Medición Nro. 1	38
Figura 50. VIF de las variables Modelo del Sistema de Medición Nro. 1.....	38
Figura 51. Residuos estudentizados vs. Valores ajustados para Modelo del Sistema de Medición Nro. 2	39
Figura 52. Diag. de caja de los residuos estudentizados para Modelo del Sist. de Medición Nro. 2	39
Figura 53. Residuos estudentizados vs. variables predictoras del Modelo del Sistema de Medición Nro. 2.....	40
Figura 54. Histograma de los residuos estudentizados del Modelo del Sistema de Medición Nro. 2	40
Figura 55. Gráfico Q-Q residuos estudentizados del Modelo del Sistema de Medición Nro. 2.....	40
Figura 56. Correlograma de las variables Modelo del Sistema de Medición Nro. 2	41
Figura 57. VIF de las variables Modelo del Sistema de Medición Nro. 2.....	41
Figura 58. Med. Nro 1 del 2017-03-16. Valores promedio de VC 16 min. y diferencia de valores ajustados.....	41
Figura 59. Med. Nro 1 del 2017-03-16. Promedios de VC 16 min. y nuevo promedio de valores ajustados.....	42
Figura 60. Med. Nro 1 del 2017-03-16. Promedios de VC 16 min. y alternativas para consolidar su nuevo valor..	42
Figura 61. Resultados. A izq.: Comparación en el tiempo. A derecha: Desvío de Allan	43
Figura 62. Resultados de la predicción	43

RESUMEN

En este trabajo se analiza la diferencia de fase entre los relojes atómicos INTI1 e INTI2. Cada uno de ellos cuenta con su propio equipo receptor y antena GPS, por lo tanto se consideran dos Sistemas de Medición independientes.

Dado que cada uno está vinculado con la red GPS de forma independiente es factible la utilización de la técnica Vista Común. Hoy en día en el Instituto Nacional de Tecnología Industrial, con dicha técnica se consiguen resultados cada 144 minutos.

Se propone una mejora a la técnica de Vista Común que permita establecer mediciones cada 16 minutos con un Desvío de Allan inferior a las especificaciones del equipo, lo cual asegura una buena estabilidad en la medición de diferencia de fase. Esta nueva metodología no solo propone una mejora en el Desvío de Allan, sino que además disminuye los tiempos de la ventana de integración aproximadamente diez veces.

INTRODUCCIÓN

El segundo es una de las siete unidades base que conforman el Sistema Internacional de Unidades. Los relojes atómicos de cesio son el patrón de tiempo para los Institutos Nacionales de Metrología. A partir de ellos se define el segundo [s] y con su inversa se define la frecuencia [1/s]. Si bien el paso del tiempo es una concepción del hombre, la misma es relevante para establecer todo tipo de redes de comunicación por ejemplo: redes de televisión, teléfono, celular, datos, etc. La generación eléctrica está íntimamente vinculada con la frecuencia, y a su vez el diseño de las líneas de transmisión eléctrica está ligado a la frecuencia de generación. Esta magnitud cobra suma importancia cuando se requiere que dispositivos como: marcapasos, respiradores artificiales y desfibriladores funcionen conforme a una determinada frecuencia. Es por esto que el período de tiempo transcurrido entre dos eventos debe ser el mismo para todos aquellos países firmantes del MRA, Acuerdo de Reconocimiento Mutuo, dentro de las incertidumbres declaradas.

En los años 70, la calibración de un reloj atómico se realizaba mediante un contador o frecuencímetro, midiendo la diferencia de fase de dos relojes atómicos in situ. Esto implicaba el traslado de los relojes entre institutos, con la complejidad agregada de mantener los instrumentos energizados durante todo el trayecto. Este método de comparación es denominado Medición Directa, ver Figura 1.

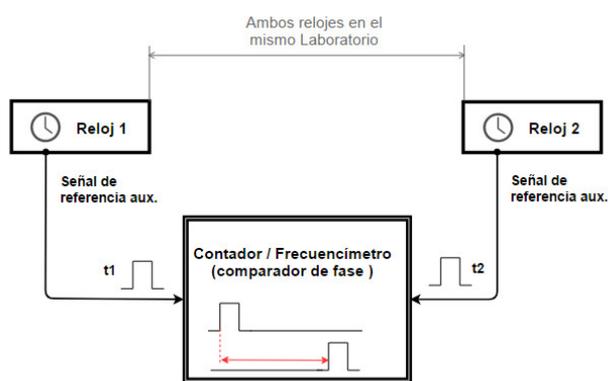


Figura 1. Esquema de Medición Directa

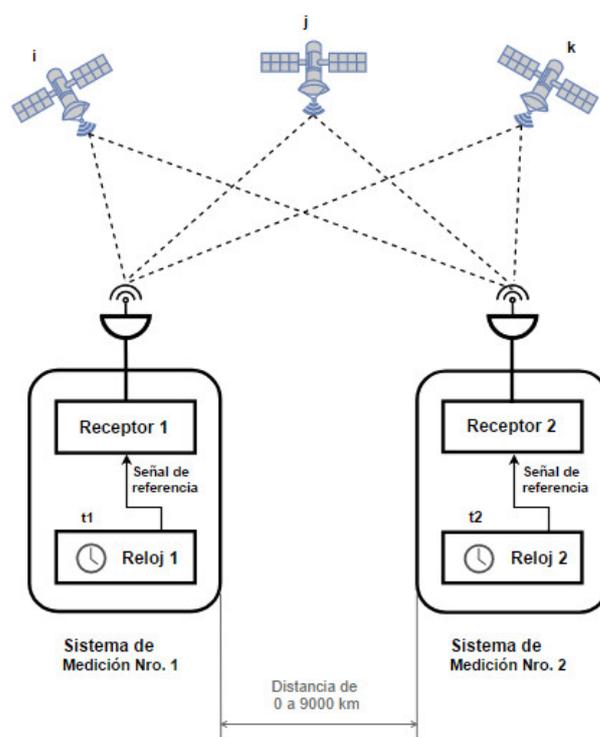


Figura 2. Esquema de Vista Común

Con la implementación de la red GPS, es posible realizar este proceso de calibración de manera más sencilla. Una de las técnicas para comparar relojes atómicos utilizando la red GPS es conocida como Vista Común [1]. La misma permite comparar relojes muy distantes entre sí, sin la necesidad de tener que trasladar alguno de los dos relojes.

El Instituto Nacional de Tecnología Industrial (INTI) utiliza ambas técnicas: Medición Directa y Vista Común para conocer la diferencia de fase y la estabilidad de sus dos relojes. Para cuantificar la estabilidad de los relojes atómicos se utiliza una medida conocida como Varianza de Allan [2].

La Varianza de Allan, es una medida de estabilidad en la frecuencia de relojes atómicos, osciladores y amplificadores. Lleva el nombre del físico David W. Allan y se expresa matemáticamente como $\sigma_y^2(\tau)$. Su objetivo es estimar la estabilidad inherente al error de proceso y no la debida a los errores de sistema, derivas de frecuencias o efectos térmicos.

La Varianza de Allan se define como la mitad del promedio de tiempo de los cuadrados de las diferencias entre las lecturas sucesivas de diferencia de fase muestreada durante el período de muestreo. Esto implica que

depende del período de tiempo entre muestras, razón por la cual se evalúa mediante un gráfico en lugar de un solo número. Una Varianza de Allan baja es una característica de un reloj con buena estabilidad durante el período de medición. Para mantener una coherencia en las unidades, se utiliza el Desvío de Allan que es la raíz cuadrada de la Varianza de Allan, $\sigma_y(\tau)$.

Con la finalidad de conseguir una estabilidad buena y aceptada entre sus pares, el INTI utiliza mediciones de Vista Común promediadas cada 144 minutos. Las mediciones de Vista Común promediadas cada 16 minutos presentan un Desvío de Allan mayor a las especificaciones del equipo y actualmente son descartadas.

En este trabajo se analizaron los datos de dos relojes atómicos: INTI1 e INTI2. Ambos cuentan con receptores y antenas GPS independientes, cada uno de ellos conforma un Sistema de Medición. Se cree que mediante un análisis exhaustivo de los datos se podría mejorar la estabilidad alcanzada hoy en día y también reducir el tiempo de la ventana de integración. Esto no solo generaría una variante en el método actual de trabajo, sino que también permitiría mejorar la incertidumbre asociada a la medición.

Para lograr esto, se analizará en detalle el protocolo de Vista Común. Una vez comprendido, mediante herramientas de Minería de Datos, se propondrán mejoras que permitan lograr un salto de calidad en la medición.

MATERIAL Y MÉTODOS

VISTA COMÚN CADA 144 MINUTOS

Los datos con los que se comparan ambos Sistemas de Medición - INTI1 e INTI2 - pueden ser descargados de <http://www.aggio-conicet.gob.ar/rnt/>. Los mismos presentan el siguiente formato:

MJD	Prom. Dif [ns]
57828,01	93,50
57828,11	92,79
57828,22	93,76
57828,32	92,14
57828,42	93,11
57828,52	93,57
57828,62	93,28
57828,72	91,38
57828,82	93,92
57828,92	94,47
57829,01	96,17

Tabla 1. Datos de VC 144 minutos

El período bajo estudio corresponde a seis días completos de medición, desde el 16 al 21 de Marzo de 2017.

Se puede apreciar que la Tabla 1 el conjunto de datos es bastante simplificado, solamente presenta dos columnas.

La primer columna denominada MJD, Fecha Juliana Modificada, es un formato de fecha generalmente utilizado por los astrónomos. Es de notar que cada medición se realiza cada 144 minutos, lo que equivale a 2,4 horas, por lo tanto por cada día bajo estudio se tendrán diez mediciones.

En la segunda columna se encuentra el promedio de la diferencia de fase entre el Reloj Atómico INTI2 y el Reloj Atómico INTI1 cada 144 minutos. La misma está expresada en unidades de tiempo, nanosegundos [ns].

Se realiza un gráfico de dispersión con los datos de la Tabla 1 y los mismos se interpolan linealmente punto a punto, se obtiene el gráfico que sigue a continuación.

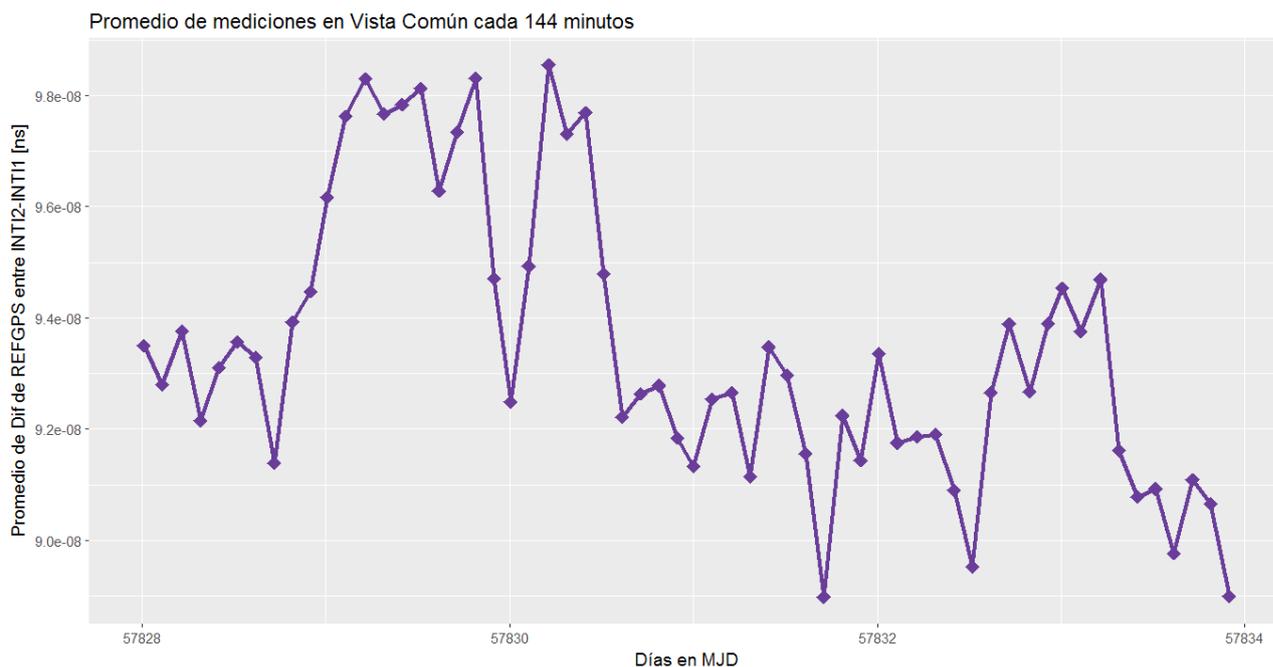


Figura 3. Diferencias de fase entre INT2 e INTI1 en VC 144 min durante seis días

Se recuerda que la segunda columna de la Tabla 1, es el resultado de varias mediciones a lo largo de 144 minutos. Dicho período entre las muestras tiene como finalidad minimizar el Desvío de Allan, el cual sirve para cuantificar la estabilidad de los relojes atómicos.

En primer medida, se puede pensar que la Figura 3 tiene muchas oscilaciones, picos y valles de notable significancia. Sin embargo, si se presta atención a la escala de la ordenada las oscilaciones están dentro de los 5 ns.

El período mostrado no fue elegido al azar, durante este período se realizó una Medición Directa, la misma consiste en conectar un frecuencímetro a los relojes atómicos INTI1 e INTI2 y medir su diferencia de fase cada 1 segundo. Ver Figura 1. Esquema de Medición Directa.

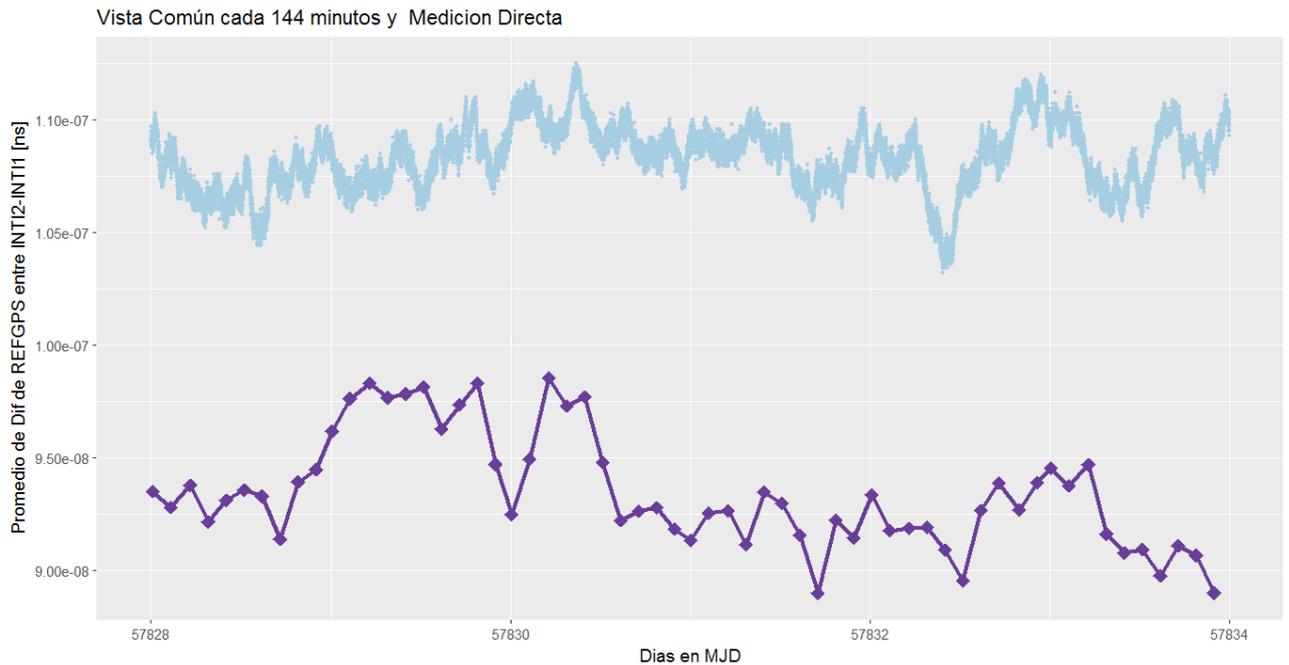


Figura 4. Comparación ente Vista Común cada 144 minutos y Medición Directa

Si se observa la Figura 4 queda de manifiesto una diferencia entre la Medición Directa - color celeste - y el método de Vista Común promediado cada 144 minutos - color violeta -. La Medición Directa, no solo es más densa sino que también esta desplazada aproximadamente 15 ns (posiblemente debido a la calibración de los cables de conexionado).

A continuación se grafica ambas señales superpuestas y se analiza el Desvío de Allan.

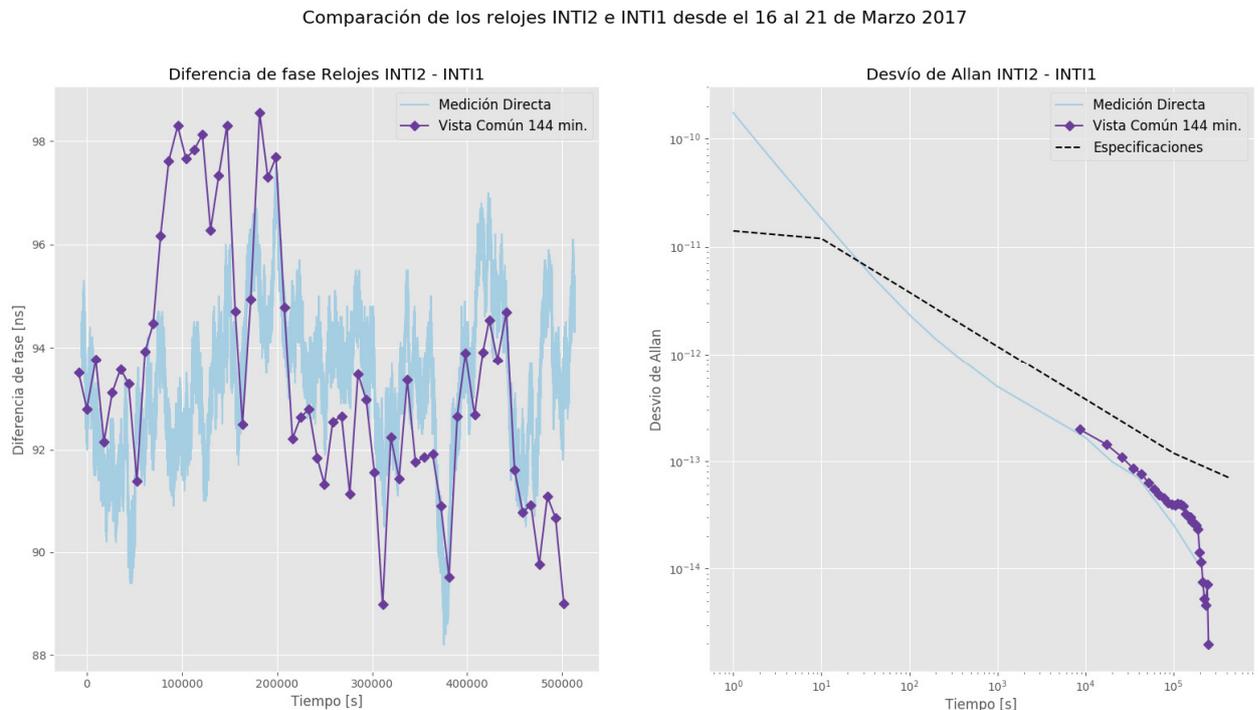


Figura 5. A izq.: VC 144 min y Med. Directa -superpuestas-. A derecha: Desvío de Allan

En la Figura 5, a la izquierda, se ha realizado lo que se denomina gráfico de control. Se superponen ambas

mediciones desplazadas y se observa que ambas señales tengan el mismo comportamiento en el tiempo. En líneas generales, el método de Vista Común sigue la forma de la nube de puntos. Sin embargo, en el segundo día de medición se observa un notable corrimiento por exceso en el método de Vista Común.

A la derecha de la Figura 5, se observa un gráfico del Desvío de Allan vs. Tiempo entre muestras. Para una simple compresión del Desvío de Allan se ha trazado una línea punteada -color negro- que permite evaluar la estabilidad de la medición. Si la medición se encuentra por debajo de la línea punteada, es aceptada y su estabilidad está dentro de las especificaciones del fabricante del reloj atómico. En color celeste se puede observar la variación del Desvío de Allan para la Medición Directa, la misma comienza desde 1 segundo ya que el período entre muestras es de 1 segundo. Conforme aumenta el período tomado entre muestras el Desvío de Allan disminuye. La medición de Vista Común - color violeta - comienza desde 144 minutos - cerca de 10^4 segundos -, y también disminuye a medida que aumenta el período entre muestras.

VISTA COMÚN CADA 16 MINUTOS

Dado que el conjunto de datos - Vista Común cada 144 minutos- contiene solamente dos columnas, se decide explorar el origen de los mismos. Un registro de una medición de Vista Común cada 144 minutos es resultado del promedio de nueve mediciones. Cada una de esas nueve mediciones, son el promedio de la diferencia de fase entre el Reloj Atómico INTI2 y el Reloj Atómico INTI1 durante 16 minutos. Si se grafica la evolución en el tiempo de dichas mediciones promedio, se observa lo siguiente:

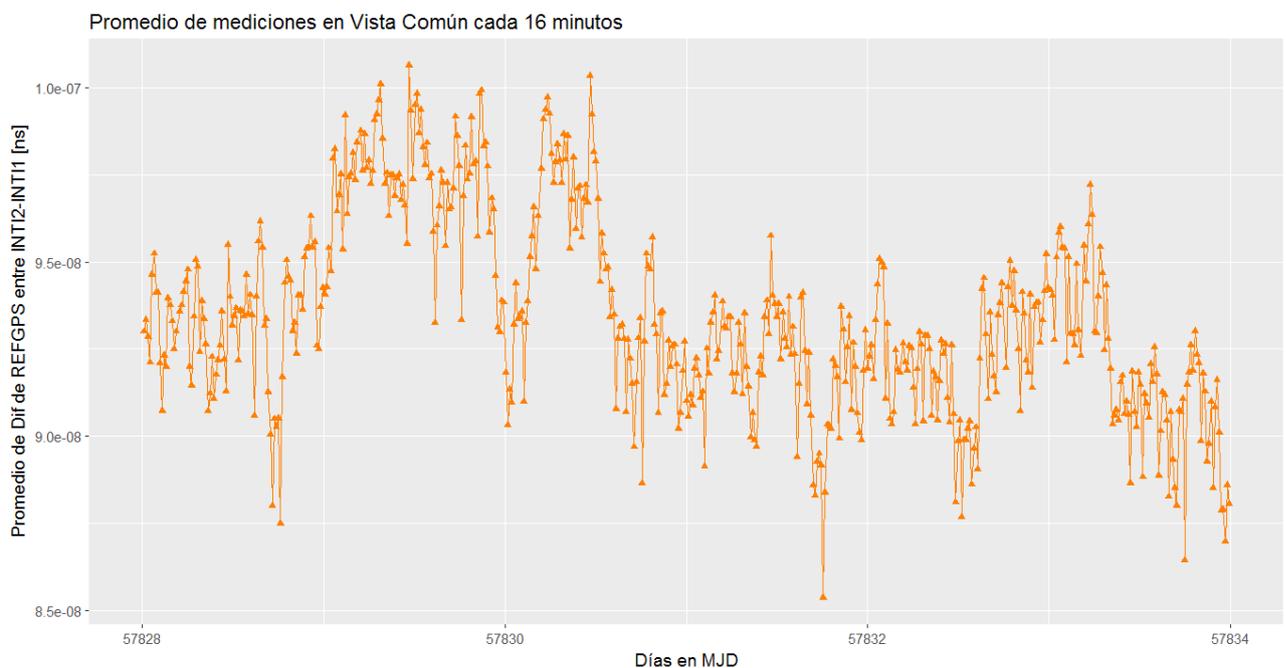


Figura 6. Evolución en el tiempo de los diferencias de fase entre INT2 e INTI1 en VC 16 minutos.

Es fácil de apreciar que la medición cada 16 minutos tiene muchas variaciones, esto es comúnmente conocido como ruido de medición. Si se evalúa el Desvío de Allan para estas mediciones, se puede comprobar que las mismas no cumplen con las especificaciones mínimas de estabilidad.

Comparación de los relojes INTI2 e INTI1 desde el 16 al 21 de Marzo 2017

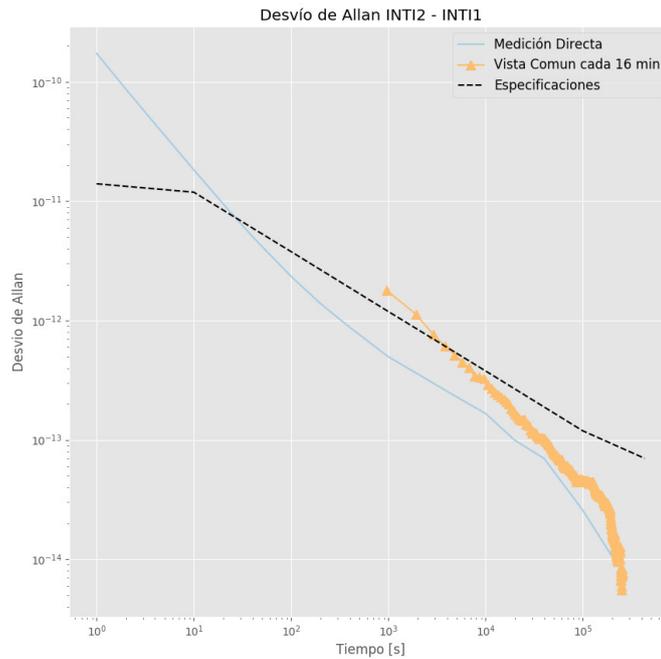


Figura 7. Desvío de Allan para Vista Común cada 16 minutos.

En la Figura 7, se puede apreciar que para 16 minutos - cerca de 10^3 segundos - la técnica de Vista Común cada 16 minutos no produce mediciones estables. Esta es una de las principales razones por la cual las mediciones promedio cada 16 minutos son descartadas.

VISTA COMÚN

En esta sección se explicará en detalle el algoritmo de Vista Común. Cada Sistema de Medición genera un archivo que contiene mediciones durante 24 horas con intervalos de 16 minutos. El primer paso, es consolidar los días bajo estudio en un gran conjunto de datos. Este conjunto de datos tendrá todas las mediciones realizadas por el Sistema de Medición INTI1. También existirá otro conjunto de datos para el Sistema de Medición INTI2, que contendrá las mediciones realizadas durante el período bajo estudio.

Por ejemplo el archivo generado por el Sistema de Medición INTI1 tiene el siguiente formato:

PRN	CL	MJD	STTIME	TRKL	ELV	AZTH	REFSV	SRSV	REFGPS	SRGPS	DSG	IOE	MDTR	SMDT	MDIO	SMDI	CK
1	FF	57828	1400	780	141	37	-514984	-1	-839	124	36	3	327	152	124	20	DA
3	FF	57828	1400	780	563	697	1068116	2	-802	-29	13	80	99	7	59	4	A4
6	FF	57828	1400	780	330	2400	-3262148	-50	-805	-53	10	33	150	-11	84	-4	E1
7	FF	57828	1400	780	397	3327	-3718750	24	-745	-6	15	62	129	-19	74	-9	FB
9	FF	57828	1400	780	520	2420	-3243654	-77	-771	18	12	47	104	-5	62	-2	DA
16	FF	57828	1400	780	210	910	-295639	0	-753	-43	21	68	227	-30	107	-6	D8

Tabla 2. Datos crudos del Sistema de Medición 1

La definición y el significado de cada una de las variables se presenta a continuación:

Categoría	Explicación
PRN	Número de satélite
CL	Clase de seguimiento en Vista Común - constante definida en el Sistema de Medición -
MJD	Fecha UTC en formato Fecha Juliana Modificada
STTIME	Hora, minutos y segundos (UTC)
TRKL	Tiempo de seguimiento en segundos
ELV	Elevación del satélite
AZTH	Azimut del satélite
REFSV	Diferencia de tiempos en [0,1 ns] entre el reloj referencia del laboratorio y el tiempo del satélite, referida al punto medio de un ajuste lineal
SRSV	Pendiente determinada en el ajuste lineal de REFSV [0,1 ps]
REFGPS	Diferencia de tiempos en [0,1 ns] entre el reloj referencia del laboratorio y el tiempo del Sistema GPS, referida al punto medio de un ajuste lineal
SRGPS	Pendiente determinada en el ajuste lineal de REFGPS [0,1 ps]
DSG	Raíz cuadrada de los residuos del ajuste lineal de REFGPS
IOE	Índice de efemérides
MDTR	Modelado de retardo troposférico [0,1 ns] referido al punto medio de un ajuste lineal
SMDT	Pendiente determinada en el ajuste lineal de MDTR [0,1 ps]
MDIO	Modelado de retardo ionosférico [0,1 ns] referido al punto medio de un ajuste lineal
SMDI	Pendiente determinada en el ajuste lineal de MDIO [0,1 ps]
MSIO ¹	Valor de retardo ionosférico medido [0,1 ns] referido al punto medio de un ajuste lineal
SMSI ¹	Pendiente determinada en el ajuste lineal de MSIO [0,1 ps]
ISG ¹	Raíz cuadrada de los residuos del ajuste lineal de MSIO
CK	Verificación de la correcta recepción de los datos en formato hexadecimal

Tabla 3. Detalle de las variables del conjunto de datos

Una vez conformados los archivos de cada Sistema de Medición, se genera un archivo de Vista Común. En dicho archivo figuran aquellos satélites que coinciden en ambos Sistemas de Medición con un intervalo de superposición de ocho minutos.

De esta forma se obtienen mediciones cada 16 minutos correspondientes a ambos Sistemas. A continuación se realiza la diferencia entre los Sistemas de Medición por satélite, quedando como resultado final la diferencia de fase entre los relojes atómicos de INTI1 e INTI2 por satélite.

$$\text{Sistema de Medición Nro. 1} \rightarrow \text{REFGPS} = \text{RelojSatélite}_x - \text{RelojINTI}_1$$

$$\text{Sistema de Medición Nro. 2} \rightarrow \text{REFGPS} = \text{Reloj Satélite}_x - \text{RelojINTI}_2$$

$$\text{Diferencia de Sistemas}_{\text{Nro1-Nro2}} = \text{RelojINTI}_2 - \text{RelojINTI}_1$$

¹ El Sistema de Medición Nro. 1 - INTI 1, no presenta estas capacidades de medición

Si bien el efecto del satélite ha sido eliminado de nuestra ecuación, si por cada intervalo de 16 minutos se tiene N satélites en Vista Común, se tendrán N Diferencias de Sistemas. La forma de evaluar la diferencia de fase para un intervalo de 16 minutos es realizar un promedio de las N Diferencias de Sistemas.

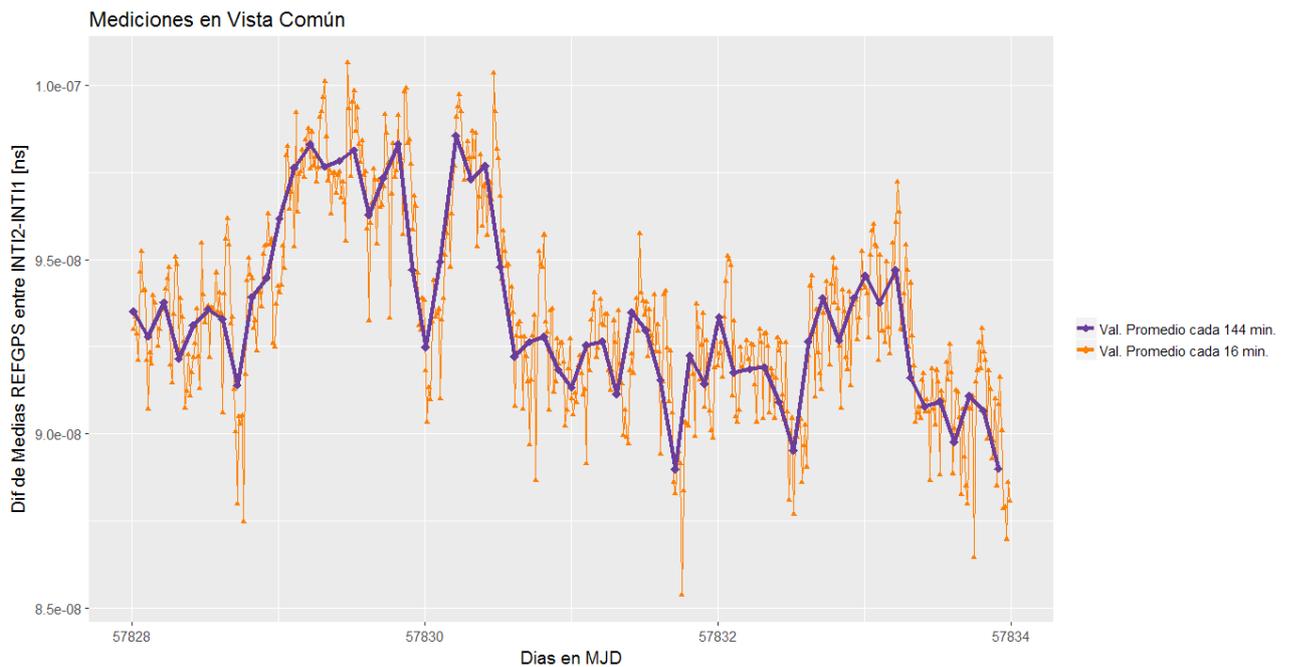


Figura 8. Comparación de VC 144min y VC 16 min.

En color naranja se puede observar el promedio de las N Diferencias de Sistemas, anteriormente fue presentado como Vista Común cada 16 minutos. En color violeta se ha superpuesto el promedio de 9 mediciones de 16 minutos, es decir Vista Común cada 144 minutos.

Vista Común 16 minutos	
Estadísticos descriptivos	Valor
Mínimo	85,37 [ns]
Primer cuartil	91,54 [ns]
Mediana	93.12 [ns]
Media	93.46 [ns]
Tercer cuartil	95.24 [ns]
Máximo	100.63 [ns]

Tabla 4. Estadísticos descriptivos VC 16 min.

Vista Común 144 minutos	
Estadísticos descriptivos	Valor
Mínimo	88,97 [ns]
Primer cuartil	91,71 [ns]
Mediana	92,88 [ns]
Media	93.46 [ns]
Tercer cuartil	94,46 [ns]
Máximo	98,55 [ns]

Tabla 5. Estadísticos descriptivos VC 144 min.

Es fácil de observar que para la señal de Vista Común cada 16 minutos - color naranja - se tiene una diferencia entre el valor máximo y mínimo de 15,26 ns. Mientras que para señal de Vista Común cada 144 minutos, la diferencia entre el valor máximo y mínimo es de 9,58 ns, lo cual representa una reducción del 37% . Esta mejora se logra simplemente realizando un promedio con mayor cantidad de mediciones, lo cual permite disminuir el Desvío de Allan a valores aceptables simplemente ampliando el período de entre muestras.

ANÁLISIS DETALLADO DE VISTA COMÚN

Con la intención de comprender las variaciones que se producen y debido a la periodicidad de los satélites, efectos de la atmósfera y forma en la que los datos son generados por los Sistemas de Medición, se analizó en profundidad un día de medición de Vista Común cada 144 minutos (el mismo contiene diez mediciones).

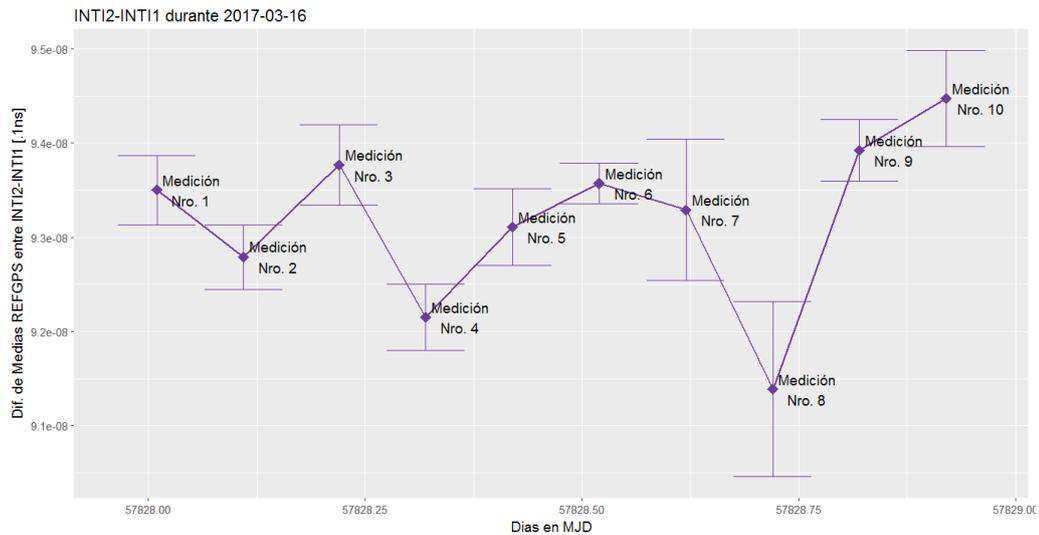


Figura 9. Día 16 de Marzo 2017 en VC 144 min.

En metrología el resultado de una medición se expresa generalmente como un valor de medida único, representado mediante el marcador, y una incertidumbre asociada² representada con las barras de error. Las barras de incertidumbre indican un entorno en donde se encuentra el valor verdadero con una probabilidad del 95%.

Cabe resaltar que varias mediciones tienen un intervalo de medición que no presenta solapamiento con su medición consecutiva. Por ejemplo: la segunda medición no concuerda con la tercer medición, lo mismo sucede con la cuarta y quinta medición. Es fácilmente apreciable que la Medición Nro. 8 tiene mayor incertidumbre que el resto. Esto deja de manifiesto que no existe uniformidad en el valor de incertidumbre a lo largo de las distintas mediciones.

Con la finalidad de conocer cómo está compuesta cada incertidumbre, se generó el siguiente gráfico.

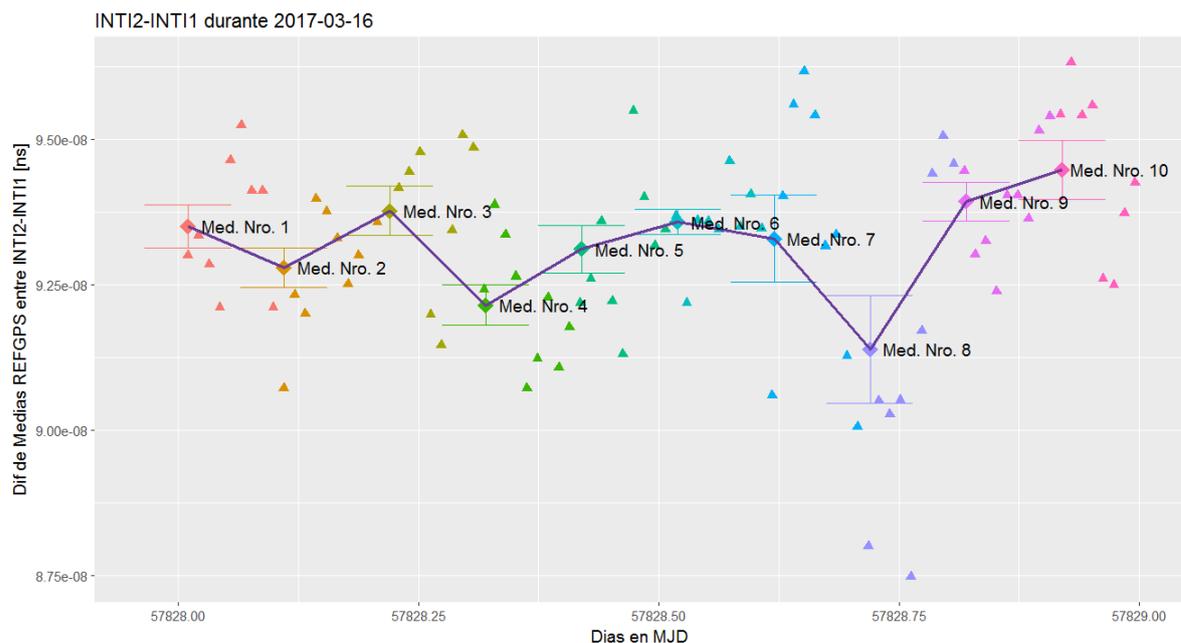


Figura 10. Día 16 de Marzo 2017 en VC 144 min. y sus medidas de VC 16 min.

² La incertidumbre asociada fue calculada como el cociente entre el desvío estándar y la raíz cuadrada de la cantidad de mediciones.

En la Figura 10 se ha identificado cada medición de Vista Común 144 minutos con un color y un marcador de forma romboidal. A su vez, se ha utilizado el mismo color y un triángulo para identificar las mediciones de Vista Común 16 minutos que con la conforman. Esto resulta clarificador para comprender la razón por la cual la Medición Nro. 8 es la que tiene mayor incertidumbre.

A continuación se observa en detalle las primeras tres mediciones de Vista Común 144 minutos para el día 16 de marzo de 2017. A cada una de las nueve mediciones de Vista Común 16 minutos, se le agregó su incertidumbre asociada en línea punteada. La misma es considerablemente mayor a la incertidumbre de cada medición romboidal.

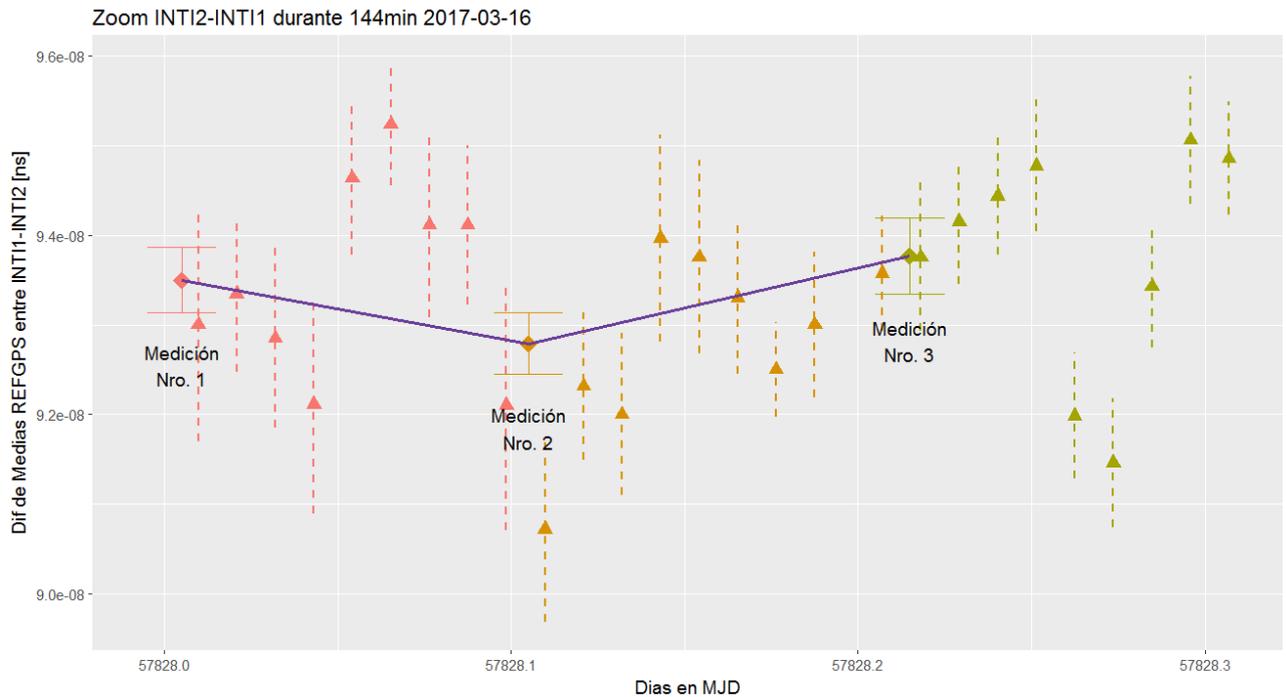


Figura 11. Detalle 2017-03-16 en VC 144 min, sus medidas de VC 16 min con incertidumbres asociadas.

En la Figura 11 se puede apreciar con facilidad el bajo valor de incertidumbre que se consigue en las tres mediciones de Vista Común 144 minutos, simplemente con realizar el promedio de las mediciones de Vista Común 16 minutos.

Cada medición de Vista Común cada 16 minutos, está conformada por los satélites en Vista Común para ambos Sistemas de Medición. Esto permitirá visualizar qué satélites están cerca del promedio y si existe algún satélite en particular que este generando estos grandes valores de incertidumbre.

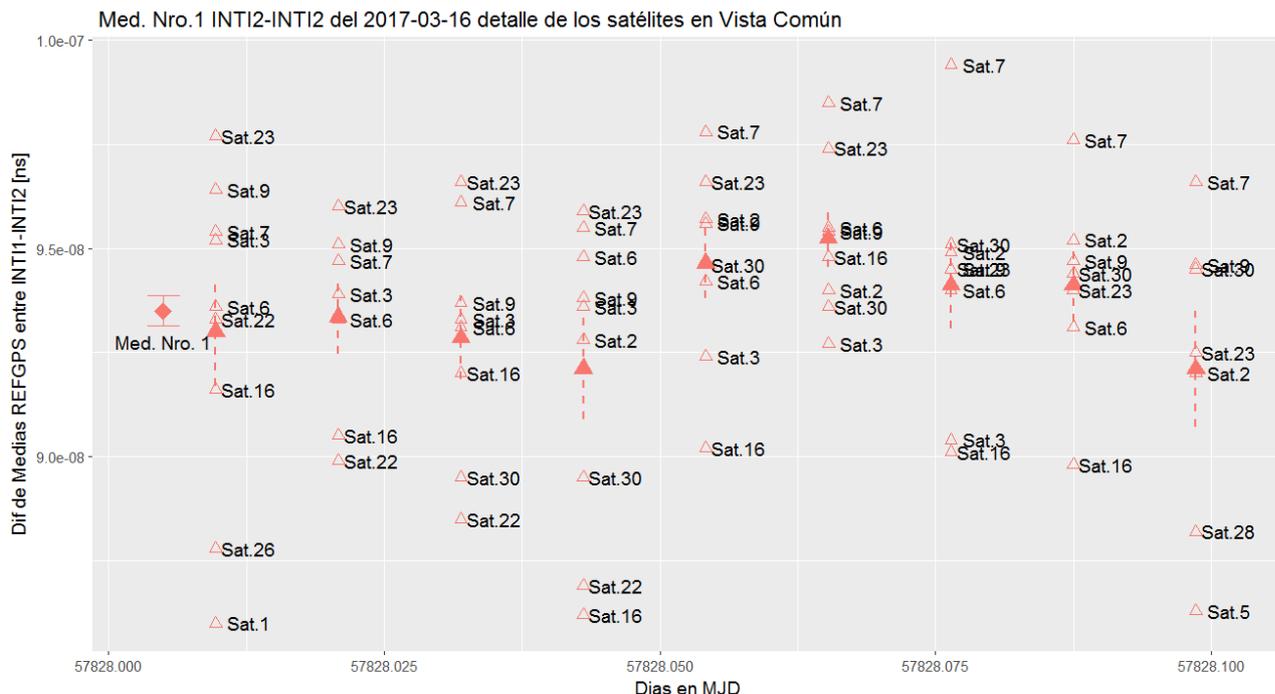


Figura 12. Med. Nro 1 del 2017-03-16. Valores promedio de VC 16 min. y Vista Común

La Figura 12, permite ver la composición de cada una de las mediciones de Vista Común cada 16 minutos para la Medición Nro. 1. Los valores promedio de las mediciones de Vista Común cada 16 minutos -representados por un triangulo lleno- y la medición obtenida de los satélites en Vista Común - representada con un triangulo vacío -.

Se puede observar la evolución que tuvo el satélite 23 entre la primer y la última medición de 16 minutos. Dicho satélite pasó de ser una medición muy alejada del centro de la distribución, a estar en el centro de la novena medición de 16 minutos. Por el contrario, el satélite 6, parece estar muy cerca del centro a lo largo de todas las mediciones. Sin embargo el satélite 22, en tan solo 64 minutos, se convierte en un medición muy alejada del centro.

Este comportamiento errático se repite con otros satélites, otras mediciones y otros días. El enfoque puntual satélite por satélite será descartado en este trabajo. El motivo de esta decisión es que el comportamiento de los satélites, esta operado y manejado por el Observatorio Naval de los Estados Unidos - USNO - y se desconoce las correcciones que realizan.

LIMPIEZA DE DATOS

Dada la naturaleza de los datos, ninguno de los Sistemas de Medición tiene datos faltantes. Esto es parte de la configuración del sistema, además cualquier dato recibido que no sea validado a través de la suma de verificación es descartado.

Esto no significa que las mediciones no contengan datos anómalos. Es por esto que la limpieza de datos es realizada en base al conjunto de datos de Vista Común, donde se conservan todas las columnas referentes a ambos Sistemas de Medición. El tamaño del conjunto datos bajo estudio, es 4214 registros y 41 variables características. Considerando que son dos sistemas separados, cada sistema será analizado por separado.

LIMPIEZA DE DATOS - SISTEMA DE MEDICIÓN 1

Una forma simple y gráfica de ver los seis estadísticos descriptivos -mínimo, primer cuartil, mediana, media, tercer cuartil y máximo - es mediante un diagrama de caja modificado³. Los diagramas de caja concentran el 50% entre los cuartiles y a partir del mismo se pueden identificar datos anómalos [3].

³ Normalmente el diagrama de caja no informa el valor medio. En este trabajo el valor medio ha sido identificado mediante un punto de color rojo.

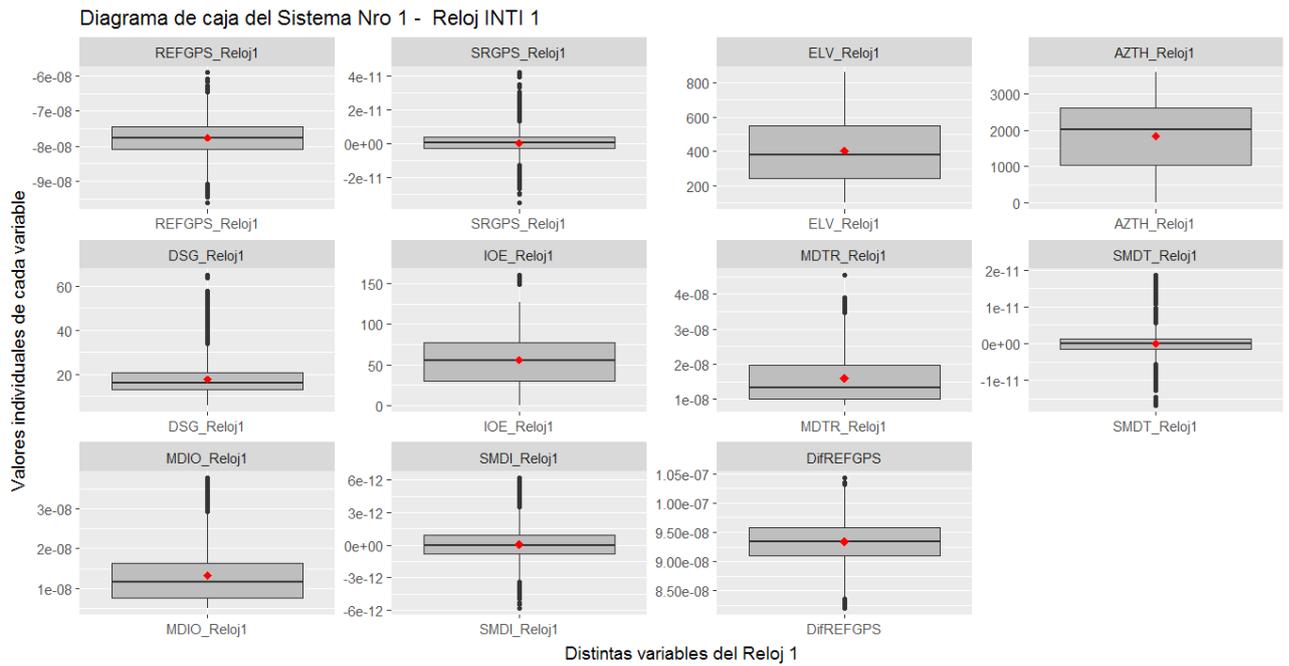


Figura 13. Diagrama de caja para el Sistema de Medición 1

Se debe prestar especial atención a la variable REFGPS_Reloj1, ya que conforma parte de la distribución de DifREFGPS. Se analiza cada uno de los diagramas de caja, en forma relativa respecto del otro. Para esto cada una de las variables de interés debe ser normalizada.

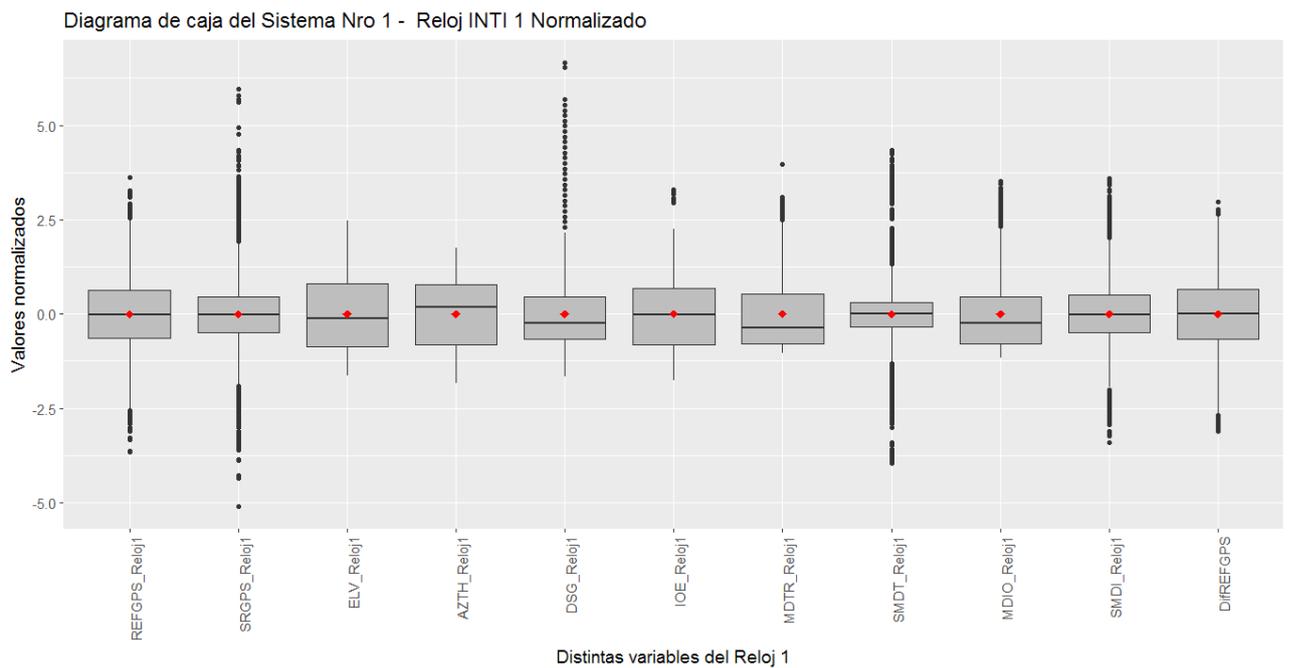


Figura 14. Diagrama de caja para el Sistema de Medición 1 - variables normalizadas -

Las variables DSG_Reloj1, IOE_Reloj1, MDTR_Reloj1 y MDIO_Reloj1 presentan datos anómalos por encima del bigote superior. En particular DSG_Reloj1 es una variable que pondera la calidad del ajuste lineal realizado por el Sistema, es una variable de sumo interés.

Dada la gran cantidad de datos bajo estudio, es normal encontrar datos anómalos en una herramienta clásica como el diagrama de caja. Por esta razón se analizan las distribuciones de cada variable mediante gráficos de densidad de kernel. En este caso se utiliza una función de kernel tipo Epanechnikov, ya que realiza un trazado más suave.

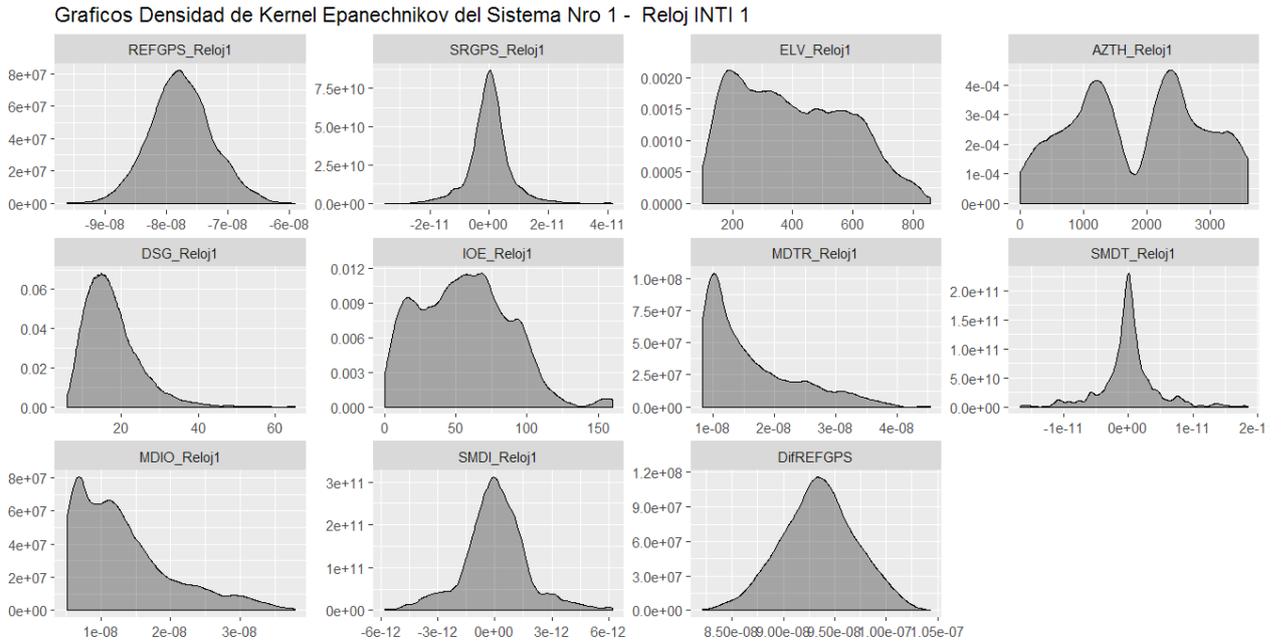


Figura 15. Gráfico de densidad de kernel para el Sistema de Medición Nro. 1

Los gráficos de densidad de kernel, evidencian las mismas conclusiones que los diagramas de caja. Es fácil de apreciar que para DSG_Reloj1, MDTR_Reloj1 y MDIO_Reloj1 existe una asimetría hacia derecha o cola hacia derecha.

Se utilizan los gráficos de kernel, en conjunto con la matriz de correlaciones para ver si existe algún patrón.

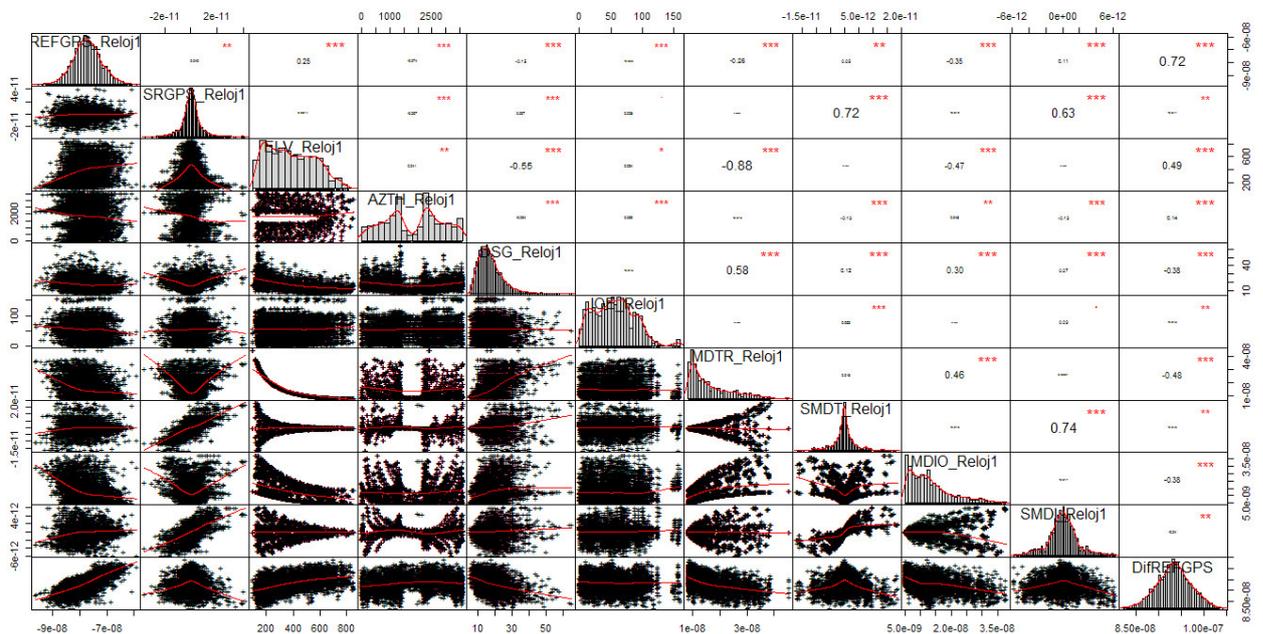


Figura 16. Extensivo correlograma para el Sistema de Medición Nro. 1

El gráfico contiene en la diagonal la distribución de las variables, debajo de la diagonal inferior gráficos de dispersión y en el triángulo superior se informan los valores de correlación. Si bien es un gráfico muy abundante en información, cuesta visualizar algún patrón.

Se realiza un correlograma con distintas intensidades de colores, con el fin de detectar visualmente correlaciones altas.

Correlograma Sistema Nro. 1

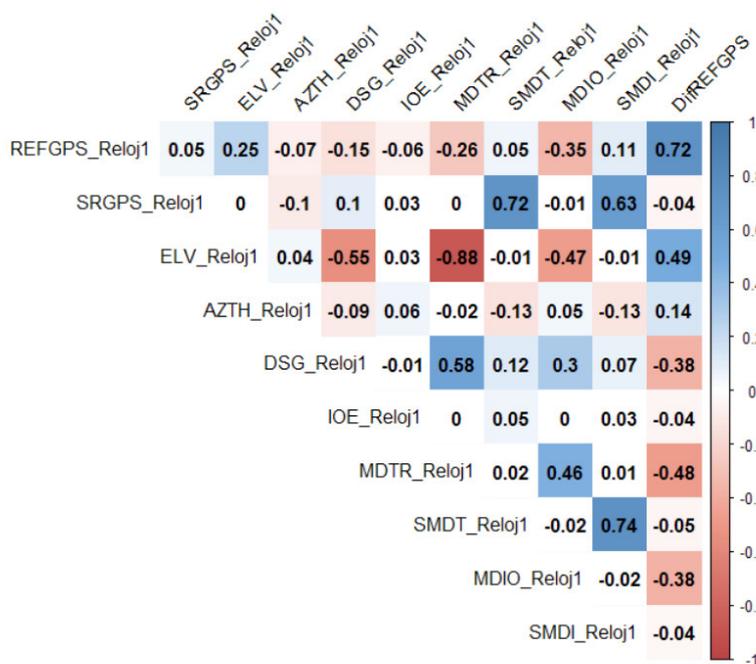


Figura 17. Correlograma para el Sistema de Medición Nro. 1

Los valores de típicos de las variables de interés son los siguientes:

	ELV_Reloj1	DSG_Reloj1	MDTR_Reloj1 [ns]	MDIO_Reloj1 [ns]
Mínimo	100	6	8,20	5,10
Primer cuartil	242	13	10	7,70
Mediana	380	17	13,4	11,50
Media	402,85	17,83	15,92	13,20
Tercer cuartil	550,75	21	19,80	16,37
Máximo	858	65	45,40	37,80

Tabla 6. Estadísticos descriptivos de variables de interés del Sist. de Medición Nro1

En base a las correlaciones Figura 17 y en base a los 6 estadísticos descriptivos Tabla 6, se definirán puntos de corte para hacer la limpieza de los datos atípicos.

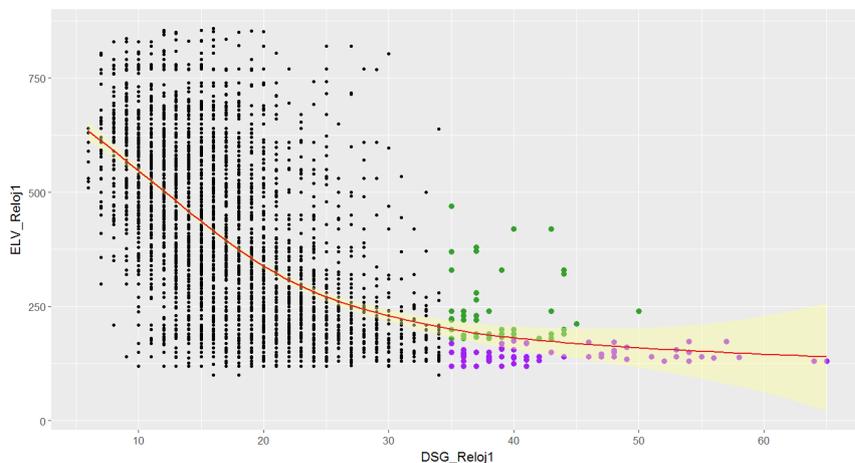


Figura 18. ELV_Reloj1 vs. DSG_Reloj1

Las variables ELV_Reloj1 y DSG_Reloj1 tienen una correlación de -0.55, es decir a medida que error de ajuste DSG_Reloj1 aumenta, la elevación disminuye.

En el gráfico se han resaltado con color, aquellos puntos que generan un valor DSG_Reloj1 mayor o igual a 35, que es donde se amplía el intervalo de validez de la regresión.

Se generó un subgrupo para aquellos puntos que presentan un valor de ELV_Reloj1 menor o igual a 175.

En este caso, ELV_Relej1 vs. MDTR_Relej1 a simple vista aparece un solo punto anómalo.

Sin embargo, explorando el conjunto de datos ese único punto son 5 (cinco) registros con el mismo valor MDTR_Relej1 = 45,4 ns - Ver Tabla 7 -.

Dicho valor coincide con el valor máximo de la variable en cuestión. Esos registros serán considerados como datos atípicos.

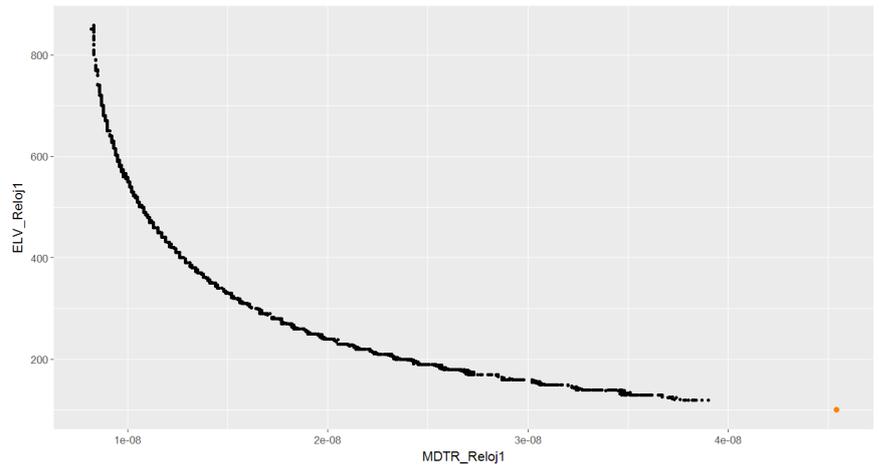


Figura 19. ELV_Relej1 vs. MDTR_Relej1

REFGPS Relej1	SRGPS Relej1	ELV Relej1	AZTH Relej1	DSG Relej1	IOE Relej1	MDTR Relej1	SMDT Relej1	MDIO Relej1	SMDI Relej1	Dif. REFGPS
-7,18E-08	1,30E-12	100	2399	16	19	4,54E-08	1,00E-13	1,36E-08	1,00E-13	9,45E-08
-7,32E-08	2,20E-12	100	2400	18	46	4,54E-08	1,00E-13	1,36E-08	1,00E-13	9,54E-08
-7,82E-08	2,60E-12	100	2400	18	44	4,54E-08	1,00E-13	1,36E-08	1,00E-13	9,15E-08
-8,01E-08	1,46E-11	100	2400	34	12	4,54E-08	1,00E-13	1,36E-08	1,00E-13	8,70E-08
-8,17E-08	-2,40E-12	100	2400	16	57	4,54E-08	1,00E-13	1,36E-08	1,00E-13	8,94E-08

Tabla 7. Detalle de los datos con MDTR = 45,5 ns

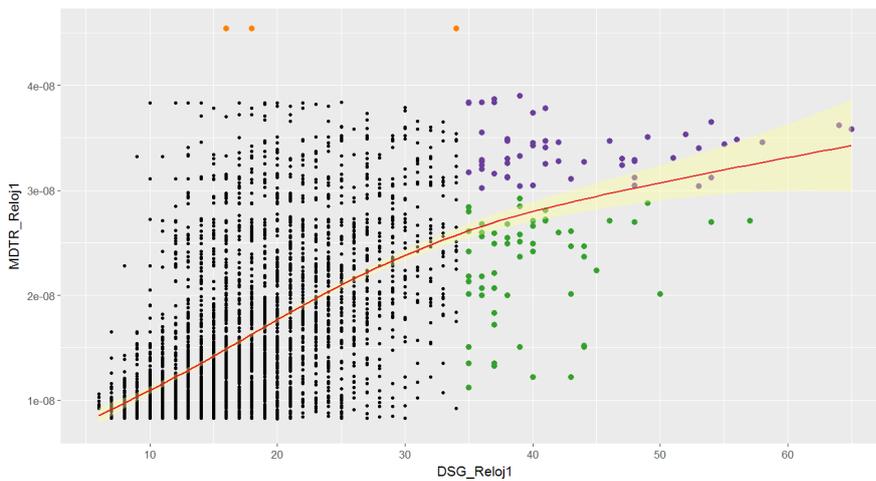


Figura 20. MDTR_Relej1 vs. DSG_Relej1

Se observa que en la Figura 20 se analiza una correlación positiva 0,58 entre la variable MDTR_Relej1 y DSG_Relej1.

Se definió un límite de 30 ns para la variable MDTR_Relej1 y se mantiene la cota superior de DSG mayor igual a 35.

Pero es de apreciar que la relación no es directa ya que los 4 valores, denotados en color naranja, con MDTR_Relej1 valor igual a 45,4 ns presentan un valor de DSG_Relej1 menor a 35.

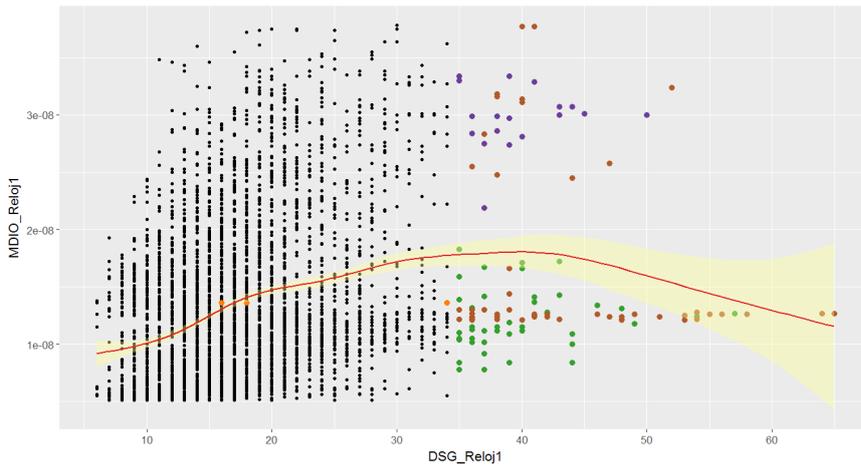


Figura 21. MDIO_Reloj2 vs. DSG_Reloj1

En la Figura 21, los puntos color violeta son valores superiores al tercer cuartil de la variable MDIO_Reloj1.

Se recuerda que la correlación entre las variables MDIO_Reloj1 y MDTR_Reloj1 es de 0,46. Esto se aprecia en el gráfico ya que son pocos los puntos con un valor de MDTR_Reloj1 mayor igual a 30 ns - color marrón -. Finalmente se concluye que los datos anómalos en la variable DSG_Reloj1 no dependen de la variable MDIO_Reloj1.

En base a esta inspección visual de datos anómalos, se encontraron 121 registros. Se decidió eliminarlos ya que representan un porcentaje muy pequeño del conjunto de datos.

LIMPIEZA DE DATOS - SISTEMA DE MEDICIÓN 2

En base a que los Sistemas son similares, y los datos son generados de la misma forma. Se utiliza la misma metodología de limpieza de datos anteriormente aplicada.

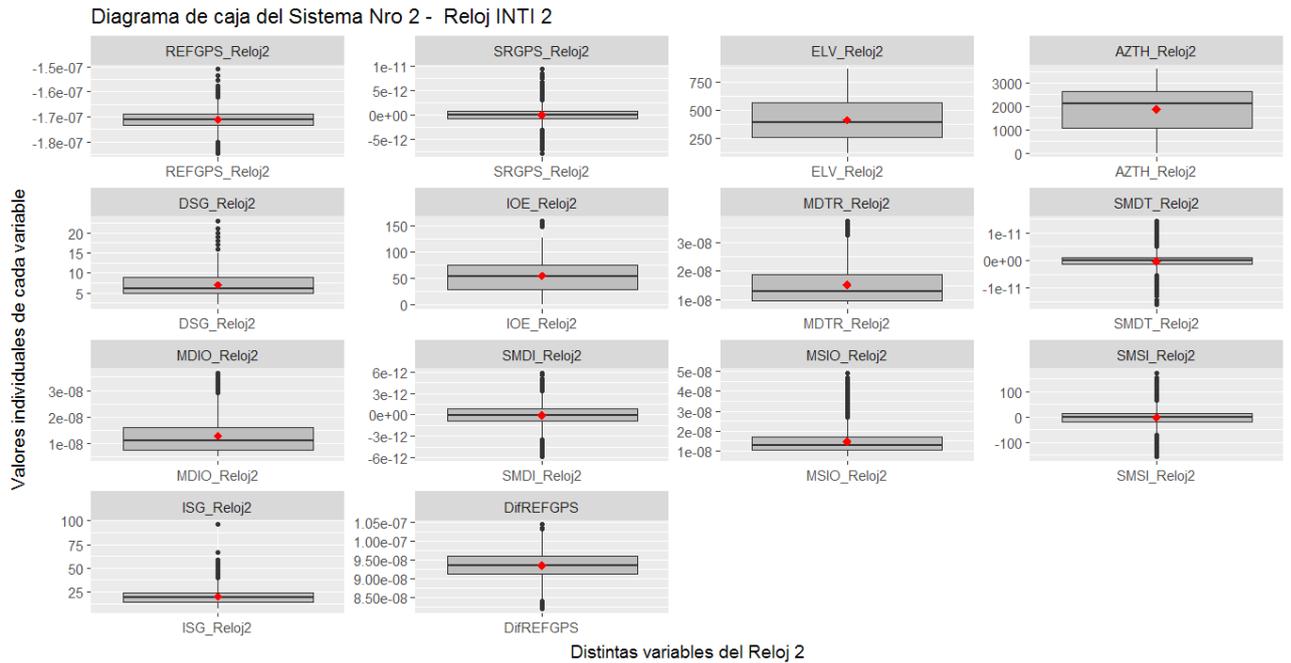


Figura 22. Diagrama de caja para el Sistema de Medición Nro. 2

Se normalizan todas las variables de interés y se ponderan los diagramas de cajas uno respecto del otro.

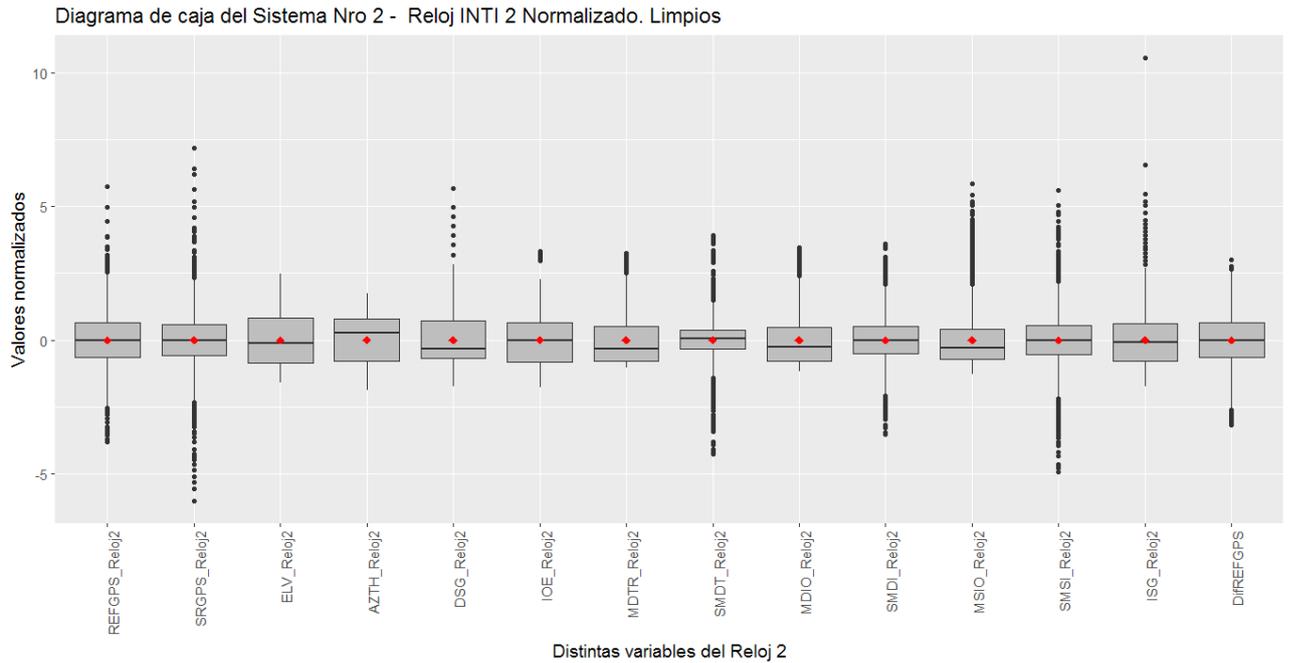


Figura 23. Diagrama de caja para el Sistema de Medición 2 - variables normalizadas -

Se observa una mayor cantidad de variables ya que el Sistema de Medición 2 permite la medición del retardo ionosférico mediante un ajuste lineal.

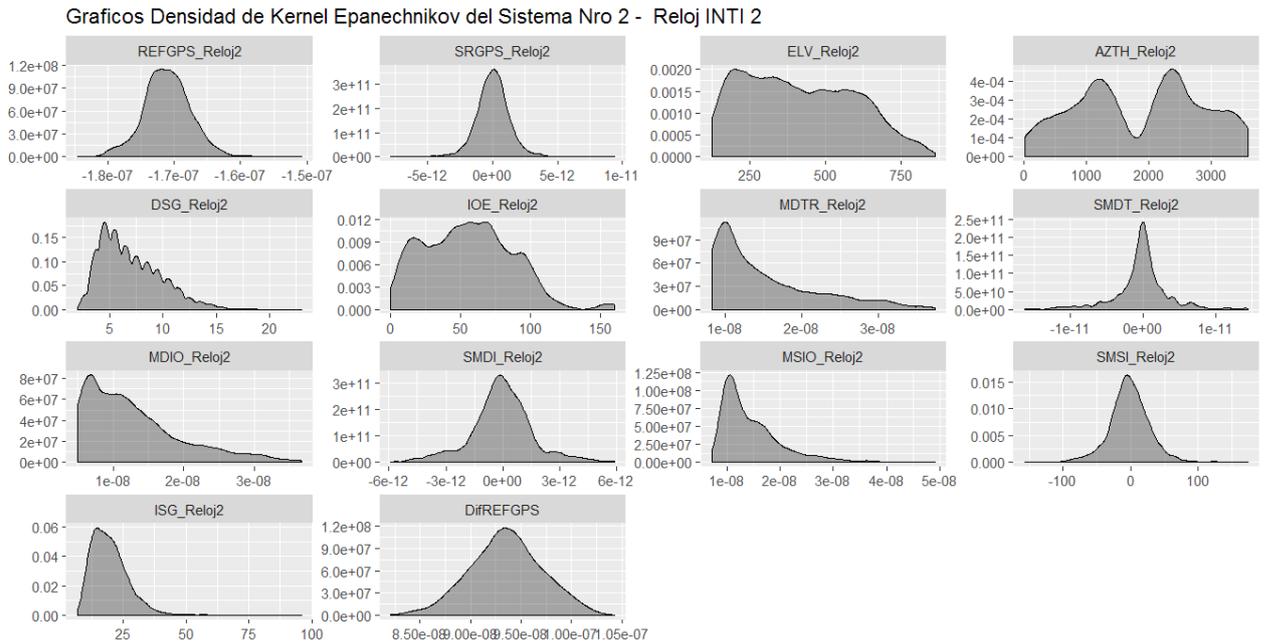


Figura 24. Gráfico de densidad de kernel para el Sistema de Medición Nro. 2

Llama la atención el rizado descendente de la variable DSG_Reloj2, esto puede deberse a la forma en la que integra punto a punto la función de Epanechnikov.

Se vuelven repetir las variables MDTR, MDIO y MSIO. En particular la variable ISG_Reloj2 claramente tiene datos anómalos.

Correlograma Sistema Nro. 2

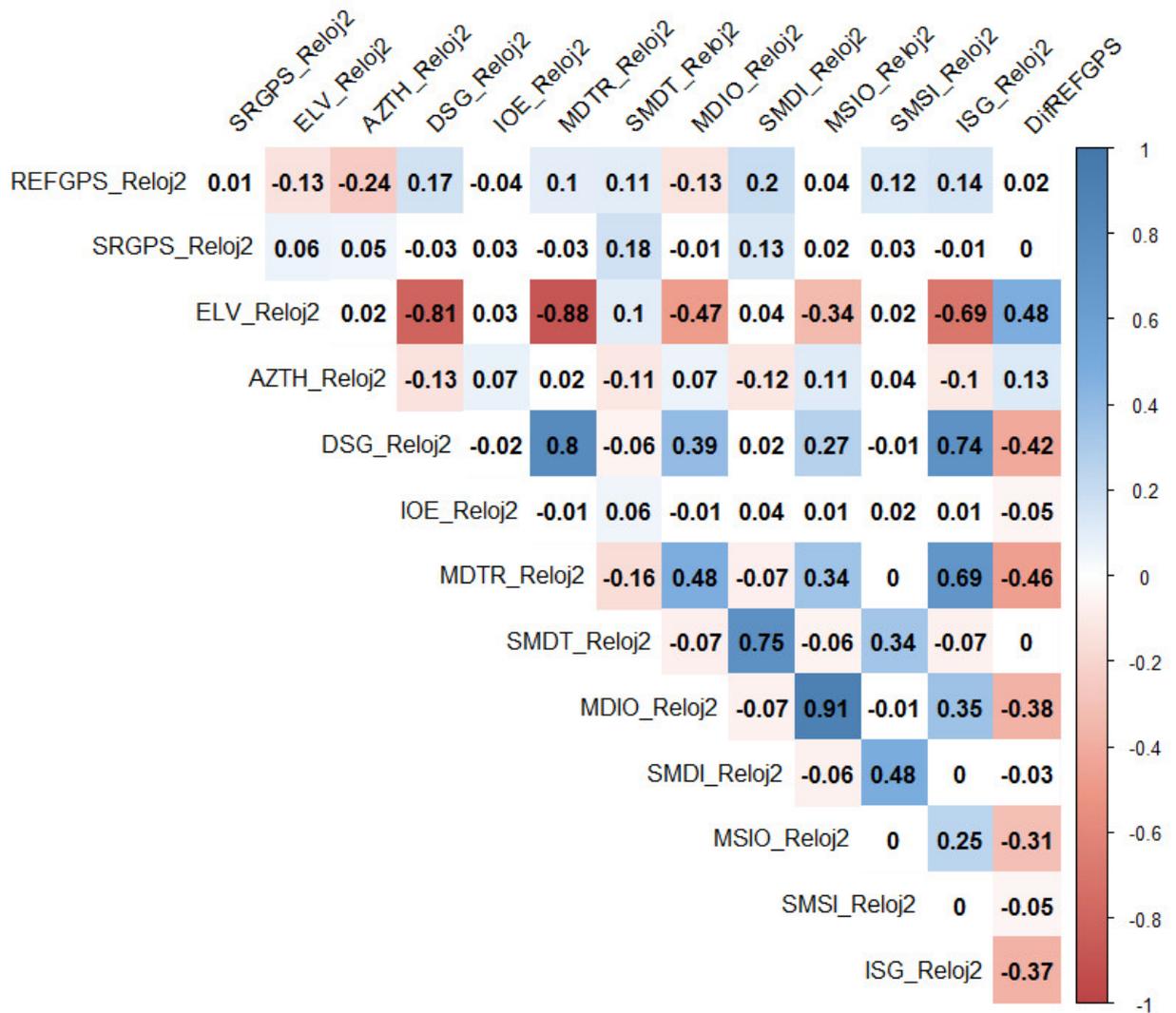


Figura 25. Correlograma para el Sistema de Medición Nro. 2

Se detallan algunos estadísticos descriptivos de las variables con alta correlación y que agregan retardo a la medición.

	ELV_Relej2	DSG_Relej2	MDTR_Relej2 [ns]	MDIO_Relej2 [ns]	MSIO_Relej2 [ns]	ISG_Relej2
Mínimo	123	2	8,20	5,00	7,10	7
Primer cuartil	257	5	9,80	7,50	10,50	14
Mediana	393	6	12,90	11,20	12,80	19
Media	414,85	6,95	15,25	12,94	14,65	19,62
Tercer cuartil	563	9	18,80	16,20	17,10	24
Máximo	864	23	37,40	36,70	49,20	96

Tabla 8. Estadísticos descriptivos de variables de interés del Sist. de Medición Nro. 2

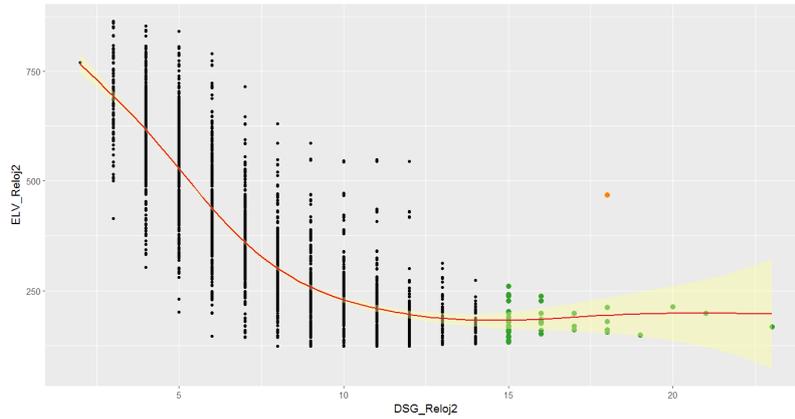


Figura 26. ELV_Relej2 vs. DSG_Relej2

La correlación ente la ELV_Relej2 y DSG_Relej2 es de -0.81. La misma es lineal hasta un valor cercano a DSG_Relej2 igual a 10.

Sin embargo, en base al estudio de los estadísticos descriptivos y los gráficos de densidad de kernel, se ha establecido un punto de corte DSG_Relej2 mayor igual a 15.

En este caso el punto anómalo en color naranja corresponde solamente a un registro.

En la Figura 27 se observan en color azul, unos registros con bajo valor de retardo, pero con un muy elevado valor de ELV_Relej2. Si se exploran los datos, se pueden encontrar 30 (treinta) registros con valores de ELV_Relej2 mayor o igual a 820. Sin embargo dichos puntos no presentan valores significativamente grandes de DSG_Relej2.

Algo similar sucede para valores superiores a 33 ns de retardo en la variable MDTR_Relej2. Estos puntos, sí presentan valores por encima del tercer cuartil de DSG_Relej2.

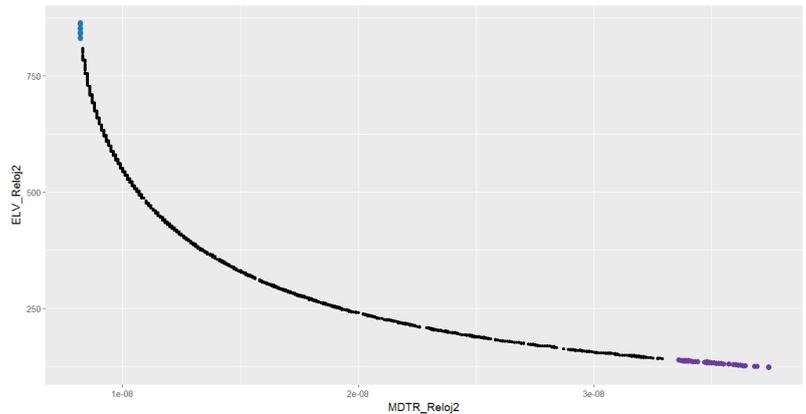


Figura 27. ELV_Relej2 vs. MDTR_Relej2

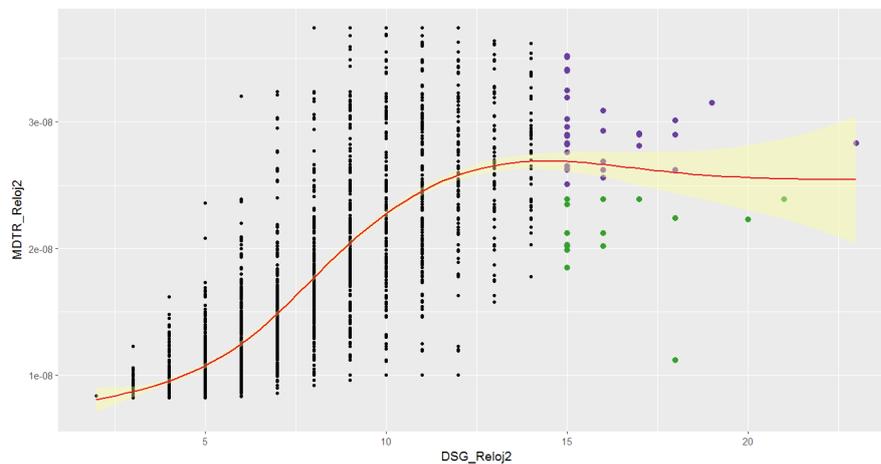


Figura 28. MDTR_Relej2 vs. DSG_Relej2

Estas dos variables MDTR_Relej2 y DSG_Relej2 presentan una alta correlación.

Se puede apreciar que a partir de un valor de DSG_Relej2 mayor o igual 15, la dispersión de puntos se abre.

Se pude subdividir en dos a ese grupo. Los puntos color violeta presentar un retardo de MDTR mayor a 25 ns.

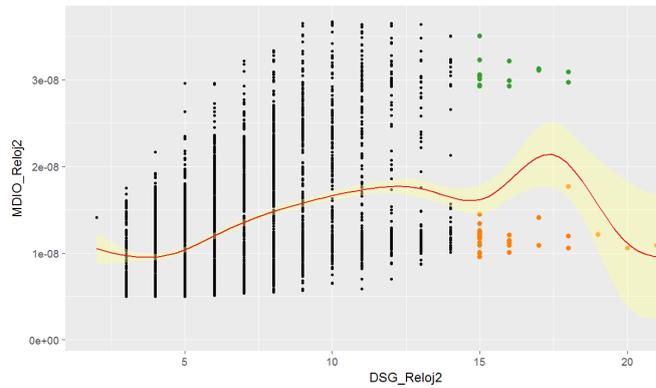


Figura 29. MDIO_Reloj2 vs. DSG_Reloj2

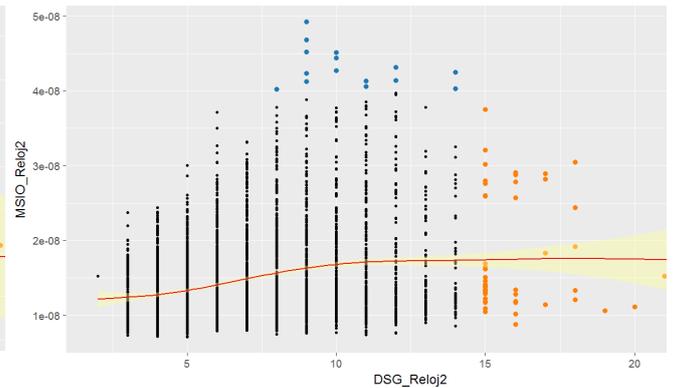


Figura 30. MSIO_Reloj2 vs DSG_Reloj2

El Sistema de Medición Nro. 2 al igual que en el Sistema de Medición Nro. 1 permite cuantificar el retardo ionosférico mediante un modelo matemático – Figura 29 -. Pero además, el Sistema de Medición Nro. 2, permite medir dicho retardo mediante la variable MSIO_Reloj2 – Figura 30 -.

Esto corrobora que el modelo matemático es congruente con la realidad ionosférica. Esto se refuerza si se establece un punto de corte para valores de DSG_Reloj2 mayor o igual 15.

Para asegurar aún más la semejanza entre las mediciones, se decide eliminar aquellos puntos que superan un retardo MSIO_Reloj2 mayor o igual a 40 ns.

Se grafica diagrama de cajas de la variable ISG_Reloj2, junto con su densidad de kernel de forma individual.

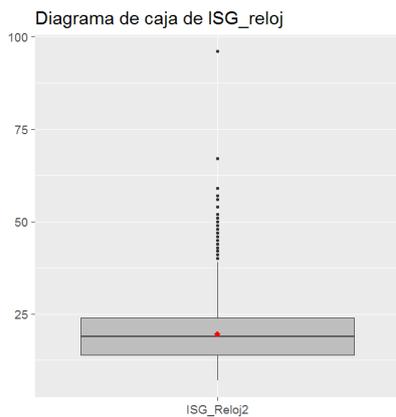


Figura 31. Diagrama de caja de ISG_Reloj2

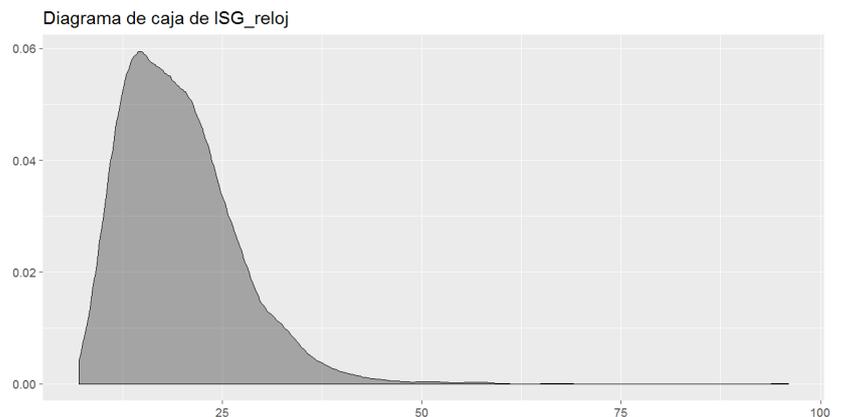


Figura 32. Densidad de Kernel de ISG_Reloj2

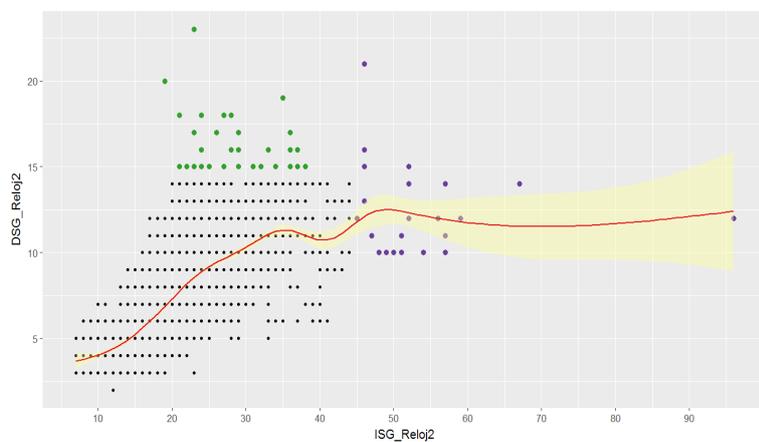


Figura 33. DSG_Reloj2 vs ISG_Reloj2

El diagrama de caja y el gráfico de densidad de kernel de ISG_Reloj2, manifiestan algunos puntos anómalos.

La correlación entre DSG_Reloj2 e ISG_Reloj2 es de 0,74.

Se puede apreciar que la regresión, extiende su intervalo de confianza para valores de ISG_Reloj2 mayores a 45. Dichos registros han sido identificados con color violeta. Así mismo se marcaron con color verde aquellos puntos con DSG_Reloj2 mayor o igual a 15. Son pocos los registros que cumplen ambas condiciones.

En base lo a lo analizado gráficamente se ha decidido eliminar dichos puntos anómalos.

MAD UNA MEDIDA DE DISPERSIÓN ROBUSTA

Un criterio para la limpieza de datos, suele ser el desvío estándar. Sin embargo, no es recomendado para determinar la estabilidad en frecuencia [4] [5]. Existen otras medidas que permiten cuantificar la dispersión de los datos de manera más robusta.

Una de esas medidas de dispersión robusta puede ser MAD – desvío absoluto de la mediana -.

$$MAD = b * Mediana_i(|x_i - Mediana_j(x_n)|)$$

Donde la constante b es igual a 1,4826 cuando se trabaja con datos que siguen una distribución normal. En el caso contrario, b se puede calcular como la inversa del tercer cuartil.

Dada naturaleza en la que son generados nuestros datos, lo más conveniente es aplicar MAD para cada valor diario del Sistema de Medición Nro. 1, Sistema de Medición Nro. 2 y finalmente para la diferencia de fase entre ambos Sistemas.

Antes de realizar la limpieza, se analiza el valor de constante b. Es decir se estudia la normalidad para cada variable y para cada día en particular. La misma fue estudiada mediante tres test de normalidad diferentes. En la tabla que sigue figuran los tres “P-valor” para cada test.

MDJday	Variable Reloj	P-valor Shapiro	P-valor Anderson	P-valor Agostino
57828	REFGPS_Reloj1	3,54E-02	6,04E-02	4,02E-01
57828	REFGPS_Reloj2	1,18E-03	3,27E-05	1,50E-02
57828	Dif_Reloj1_Reloj2	1,04E-06	5,29E-08	3,21E-06
57829	REFGPS_Reloj1	2,07E-01	1,29E-01	4,26E-01
57829	REFGPS_Reloj2	1,29E-02	3,01E-02	7,38E-02
57829	Dif_Reloj1_Reloj2	4,62E-09	1,00E-10	5,89E-10
57830	REFGPS_Reloj1	9,04E-04	5,18E-04	3,40E-04
57830	REFGPS_Reloj2	2,10E-04	4,30E-05	6,89E-03
57830	Dif_Reloj1_Reloj2	1,62E-02	2,73E-02	4,92E-02
57831	REFGPS_Reloj1	2,83E-04	3,78E-04	5,08E-04
57831	REFGPS_Reloj2	2,58E-06	1,75E-06	6,30E-05
57831	Dif_Reloj1_Reloj2	1,08E-06	2,94E-06	5,52E-06
57832	REFGPS_Reloj1	1,55E-04	4,13E-04	3,82E-05
57832	REFGPS_Reloj2	8,45E-02	4,26E-02	8,88E-01
57832	Dif_Reloj1_Reloj2	3,84E-04	1,60E-04	1,84E-04
57833	REFGPS_Reloj1	3,15E-03	2,93E-03	2,14E-01
57833	REFGPS_Reloj2	1,57E-10	2,21E-11	1,44E-10
57833	Dif_Reloj1_Reloj2	6,98E-02	1,18E-01	4,55E-02
Todos los días	REFGPS_Reloj1	4,34E-15	1,32E-19	4,55E-02
Todos los días	REFGPS_Reloj2	3,08E-06	1,31E-07	1,13E-02
Todos los días	Dif_Reloj1_Reloj2	3,24E-11	6,91E-09	1,06E-04

Tabla 9. Tabla con resultados de los test de normalidad

Los resultados de las pruebas realizadas, indican que hay suficiente evidencia estadística para asegurar la normalidad en dichas variables. Por este motivo se considera un valor de constante b = 1,4826. Se estable como umbral de corte 2 MAD respecto de la mediana – se recuerda que la mediana también es una medida robusta -.

A continuación se analizan los gráficos de densidad de kernel, antes y después de la limpieza.

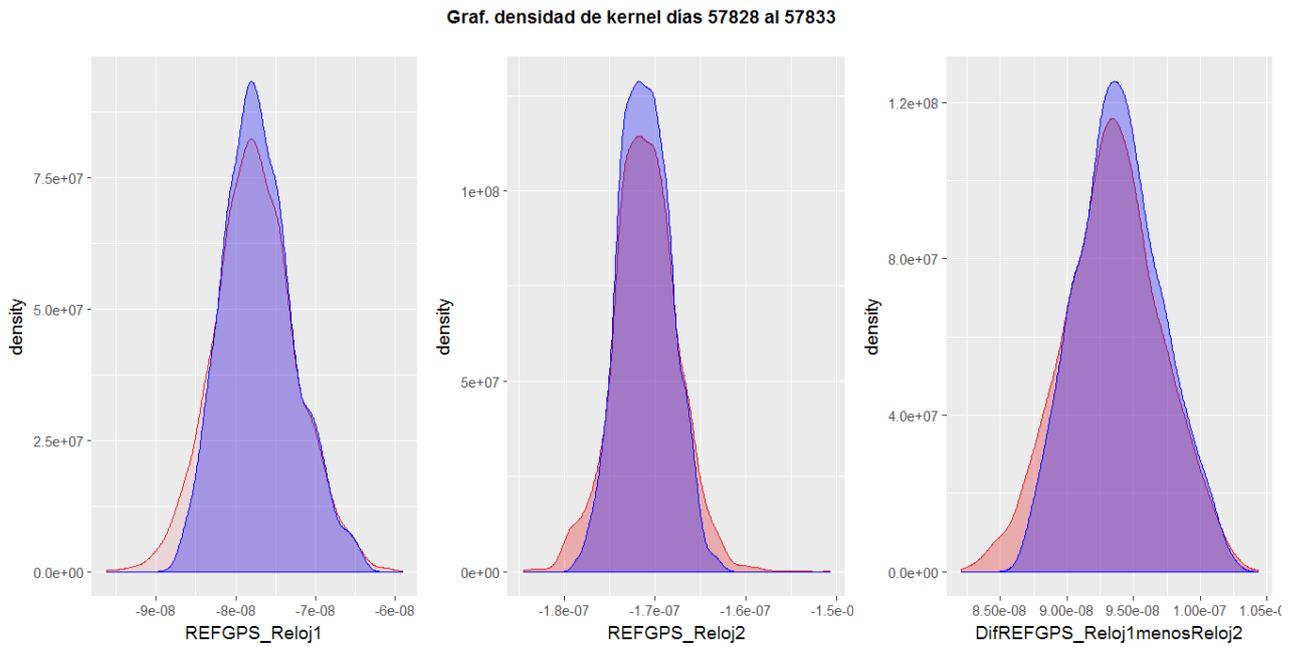


Figura 34. Densidad de Kernel, antes y después de la limpieza- 6 días-

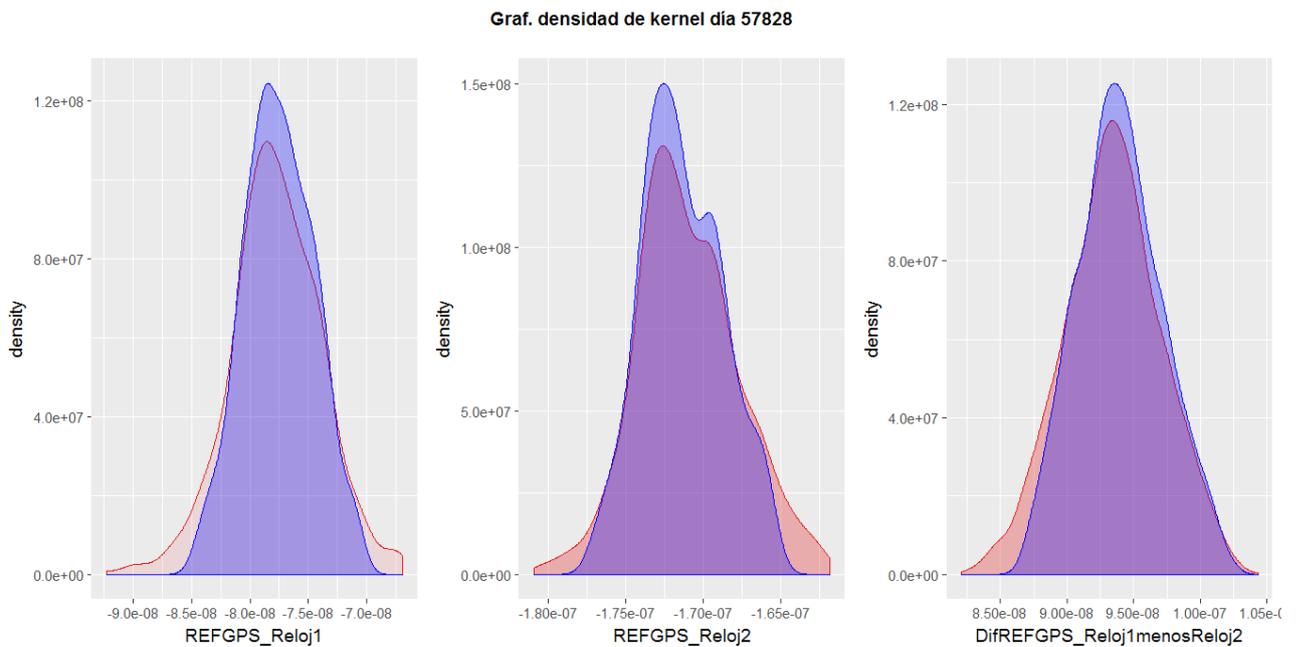


Figura 35. Densidad de Kernel, antes y después de la limpieza- 16 marzo 2017-

Es apreciable la reducción de la dispersión de valores en todas las variables. El conjunto de datos final ha sido reducido a 3416 registros.

GENERANDO UN PROMEDIO PONDERADO

Una regresión lineal, tiene el siguiente formato:

$$y = \beta_0 + \beta_1x + \varepsilon$$

La misma tiene por objetivo generar una recta, en base a unos valores observados. En general, la recta se obtiene mediante una técnica denominada mínimos cuadrados. Dicha técnica tiene por objetivo minimizar los residuos, es decir minimizar la distancia entre los valores observados y los valores ajustados.

El concepto de distancia, puede ser extrapolado a mayores dimensiones, y esto origina que el mismo principio mínimos cuadrados pueda ser aplicado en regresiones lineales múltiples [6].

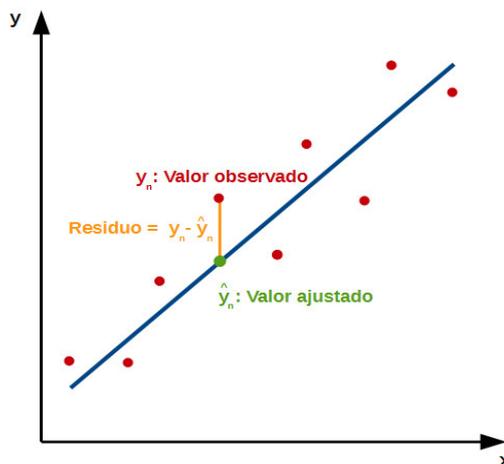


Figura 36. Regresión lineal simple

Hasta hoy en día, las mediciones de Vista Común casa 16 minutos son promediadas para generar datos que son denominados Vista Común de 144 minutos. Al realizar este segundo promedio se deprime la varianza de la medición y el Desvío de Allan alcanza la estabilidad requerida.

Este trabajo plantea la posibilidad de no descartar las mediciones de Vista Común cada 16 minutos y generarlas mediante un promedio ponderado normalizado. Un promedio ponderado normalizado, se basa en multiplicar a cada una de las variables por un determinado coeficiente. Dado que el promedio ponderado es normalizado la suma de los coeficientes debe dar uno.

En una Regresión Lineal Múltiple, se multiplica a las variables del modelo por unos coeficientes denominados betas. Los betas son los factores de ponderación que multiplican a las variables - con excepción del β_0 -

$$Regresion\ Lineal\ Múltiple \rightarrow y = \beta_0 + \beta_1x_{i1} + \beta_2x_{i2} + \dots + \beta_px_{ip} + \varepsilon \rightarrow \text{para } i = 1,2, \dots, n \\ \text{con } p \text{ variables}$$

Así como se realizó una limpieza día por día mediante MAD, se opta por el mismo criterio y se elije realizar una regresión lineal con los datos de un día.. Sin embargo, se debe recordar que el resultado de una medición de Vista Común 16 minutos es el promedio de la diferencia de fase de todos los satélites presentes en ambos sistemas dentro de un intervalo de 8 minutos.

Dado que se ha realizado una limpieza sobre los registros, se debe volver a calcular los valores promedios cada 16 minutos. Los mismos serán la variable Prom_DifREFGPS_Reloj1menosReloj2, la cual será la variable predicha de nuestro modelo de regresión lineal múltiple.

Se muestra el conjunto de variables de interés se presentan a continuación:

Variables de interés para la Reg. Lineal Múltiple		
REFGPS_Reloj1	REFGPS_Reloj2	SRGPS_Reloj1
SRGPS_Reloj2	ELV_Reloj1	ELV_Reloj2
AZTH_Reloj1	AZTH_Reloj2	DSG_Reloj1
IOE_Reloj1	IOE_Reloj2	DSG_Reloj2
MDTR_Reloj2	SMDT_Reloj1	MDTR_Reloj1
MDIO_Reloj1	MDIO_Reloj2	SMDT_Reloj2
SMDI_Reloj2	MSIO_Reloj2	SMDI_Reloj1
ISG_Reloj2	MJDdecimal_Regresion	SMSI_Reloj2
Prom_DifREFGPS_Reloj1menosReloj2		

Tabla 10. Variables de interés para la Reg. Lineal Múltiple

Dado que nuestro modelo es para 1 (un) día de medición, la variable que maneja la evolución del tiempo será una variable numérica con rango de [0 ; 1) - por esto su nuevo nombre es MJDdecimal_Regresion-. De otra forma nuestro modelo siempre estaría sesgado por el día inicial de nuestros datos 16 de marzo de 2017 (en formato MJD 57828).

La variable predicha es Prom_DifREFGPS_Relej1menosRelej2, y en base al conocimiento de dominio de los datos se determina que nuestro modelo base debe incluir como mínimo las siguientes variables: MJDdecimal_Regresion, REFGPS_Relej1 y REFGPS_Relej2.

Vale aclarar que nuestra variable predicha, Prom_DifREFGPS_Relej1menosRelej2, puede o no coincidir con los valores promedio de Vista Común 16 minutos. Observar Tabla 11.

	Vista Común 16 minutos											
	I [ns]	II [ns]	III [ns]	IV [ns]	V [ns]	VI [ns]	VII [ns]	VIII [ns]	IX [ns]	X [ns]	XI [ns]	XII [ns]
Prom. VC 16min (original)	93,00	93,34	92,85	92,11	94,64	95,24	94,11	94,11	92,10	90,72	92,32	92,00
Prom. VC 16 min. (datos limpios)	94,74	93,34	93,47	93,70	95,20	95,24	93,90	94,83	94,04	92,10	92,70	93,23

Tabla 11. Comparación de promedios en VC 16 min.

A continuación se presenta el resumen del primer modelo, generado con el lenguaje R.

Coefficientes	Estimado	Error estándar	t-valor	Pr(> t)	
Intercepto	6,63E-08	3,00E-09	22,08	< 2,22E-16	***
MJDdecimal_Regresion	4,54E-10	1,45E-10	3,14	0,0017	**
REFGPS_Relej1	2,89E-01	1,97E+02	14,66	< 2,22E-16	***
REFGPS_Relej2	-2,90E-01	2,27E+02	-12,76	< 2,22E-16	***
Códigos de significatividad: 0 '***' 0,001 '**' 0,01 '*' 0,05 '.' 0,1 '' 1					
Error estándar de residuos: 9,81E-09 con 556 grados de libertad					
R-cuadrado múltiple: 0,3074; R-cuadrado ajustado: 0,3037					
Estadístico-F: 82,27 con 3 y 556 GL; p-valor: < 2,22E-16					

Tabla 12. Primero modelo de Regresión Lineal Múltiple.

Este primer modelo, logra explicar solamente un 30% de la variabilidad de los datos.

Se decide genera un segundo modelo que incluya todas las variables. Ver Tabla 13.

Coefficientes	Estimado	Error estándar	t-valor	Pr(> t)	
Intercepto	4,04E-08	3,40E-09	11,88	< 2,22e-16	***
REFGPS_Reloj1	5,08E-01	2,11E-02	24,09	< 2,22e-16	***
REFGPS_Reloj2	-5,41E-01	2,58E-02	-20,95	< 2,22e-16	***
SRGPS_Reloj1	8,81	8,44	1,04	0,2966	
SRGPS_Reloj2	2,76E+01	3,12E+01	0,88	0,3769	
ELV_Reloj1	-1,23E-11	7,15E-12	-1,72	0,0853	.
ELV_Reloj2	9,80E-12	7,14E-12	1,37	0,1708	
AZTH_Reloj1	-1,99E-12	2,43E-12	-0,81	0,4134	
AZTH_Reloj2	1,87E-12	2,43E-12	0,76	0,4428	
DSG_Reloj1	1,49E-11	7,94E-12	1,87	0,0608	.
DSG_Reloj2	3,38E-11	2,77E-11	1,22	0,2220	
IOE_Reloj1	-2,99E-12	5,38E-12	-0,55	0,5788	
IOE_Reloj2	1,86E-12	5,37E-12	0,34	0,7294	
MDTR_Reloj1	-1,13E-01	3,51E-01	-0,32	0,7478	
MDTR_Reloj2	1,21E-01	3,68E-01	0,32	0,7417	
SMDT_Reloj1	-5,85	1,91E+02	-0,03	0,9755	
SMDT_Reloj2	1,07E+01	2,02E+02	0,05	0,9576	
MDIO_Reloj1	5,00E-01	6,26E-01	0,79	0,4250	
MDIO_Reloj2	-4,82E-01	6,30E-01	-0,76	0,4442	
SMDI_Reloj1	-2,27E+02	4,38E+02	-0,51	0,6048	
SMDI_Reloj2	1,06E+02	4,48E+02	0,23	0,8135	
MSIO_Reloj2	1,35E-01	2,48E-02	0,54	0,5854	
SMSI_Reloj2	1,60E-12	1,42E-12	1,12	0,2617	
ISG_Reloj2	2,25E-11	7,84E-12	2,86	0,0043	**
MJDdecimal_Regresion	1,24E-10	1,78E-10	0,69	0,4864	

Códigos de significatividad: 0 '***' 0,001 '**' 0,01 '*' 0,05 '.' 0,1 '' 1
Error estándar de residuos: 7,91E-10 con 535 grados de libertad
R-cuadrado múltiple: 0,5666; R-cuadrado ajustado: 0,5472
Estadístico-F: 29,14 con 24 y 535 GL; p-valor: < 2,22E-16

Tabla 13. Segundo modelo de Regresión Lineal Múltiple.

Este segundo modelo logra explicar un 54% de la variabilidad de los datos.

Con el fin de poder conseguir un mejor valor de R²-Ajustado, se genera un modelo con todas las variables y sus interacciones tomadas de a dos.

La intención de mejorar la variabilidad explicada por nuestro modelo, tiene por objetivo poder explicar el conjunto de datos y aumentar el poder predictivo del mismo. Sin embargo un modelo que explique en mucho detalle el conjunto de datos, puede resultar perjudicial para su capacidad predictiva. Esto es conocido en la jerga como "overfitting". Ver Tabla 14.

Coefficientes	Estimado	Error estándar	t-valor	Pr(> t)	
Intercepto	1,99E-07	1,59E-07	1,25	0,2104	
REFGPS_Rejoj1	4,09	1,71	2,39	0,0173	*
REFGPS_Rejoj2	3,77E-02	1,04	0,03	0,9711	
SRGPS_Rejoj1	1,25E+03	1,07E+03	1,16	0,2441	
SRGPS_Rejoj2	8,98E+03	4,19E+03	2,14	0,0329	*
...					
SRGPS_Rejoj2:MDTR_Rejoj2	-1,29E+12	5,04E+11	-2,55	0,0112	*
SRGPS_Rejoj2:SMDT_Rejoj2	-3,77E+14	3,61E+14	-1,04	0,2973	
SRGPS_Rejoj2:MDTR_Rejoj2	-1,29E+12	5,04E+11	-2,55	0,0112	*
...					
MSIO_Rejoj2:SMSI_Rejoj2	-3,15E-03	1,32E-03	-2,39	0,0174	*
MSIO_Rejoj2:ISG_Rejoj2	-8,85E-03	8,64E-03	-1,02	0,3065	
MSIO_Rejoj2:MJDdecimal_Regresion	-4,07E-01	2,40E-01	-1,70	0,0902	.
SMSI_Rejoj2:ISG_Rejoj2	8,72E-13	4,73E-13	1,84	0,0663	.
SMSI_Rejoj2:MJDdecimal_Regresion	1,56E-12	1,10E-11	0,14	0,8873	
ISG_Rejoj2:MJDdecimal_Regresion	-6,24E-11	6,76E-11	-0,92	0,3562	
Códigos de significatividad: 0 '***' 0,001 '**' 0,01 '*' 0,05 '.' 0,1 '' 1					
Error estándar de residuos: 6,28E-10 con 259 grados de libertad					
R-cuadrado múltiple: 0,8677; R-cuadrado ajustado: 0,7145					
Estadístico-F: 5,66 con 300 y 259 GL; p-valor: < 2,22E-16					

Tabla 14. Modelo de Regresión, con interacciones tomadas de a dos.

Se pueden destacar varias cosas. La primera es que el listado de variables fue recortado ya que era demasiado largo. Las variables de nuestro modelo base, están presentes pero pierden significancia. A pesar de incluir todas las variables con sus interacciones de dos, solamente se puede explicar el 71,45%.

Si bien se logró una mejora del modelo, la dificultad del mismo aumento considerablemente. Y sobre todo hay que eliminar aquellas variables fuera del modelo base, que no sean significativas. Se probó reducir la dimensión del modelo mediante técnicas "backward", "forward" y "stepwise"; sin embargo estos métodos elegían variables donde predominaba un Sistema de Medición. Por ejemplo se elegía, la variable MSIO_Rejoj2:SMSI_Rejoj2 pero no se consideraba la variable SRGPS_Rejoj2:MDTR_Rejoj2 pero no se incluía su contraparte SRGPS_Rejoj1:MDTR_Rejoj1. Otro inconveniente es que dichas interacciones son difíciles de explicar y comprender como parte del modelado del Sistema de Medición.

En base múltiples iteraciones, se define un nuevo modelo base, que para su fácil identificación se denominará Modelo Doble Interacción. En el mismo se encuentran variables tanto del Sistema de Medición Nro. 1, como del Sistema de Medición Nro. 2. Las variables incluidas en el Modelo de Doble Interacción serán:

- MJDdecimal_Regresion
- REFGPS_Rejoj1
- REFGPS_Rejoj2
- ELV_Rejoj1
- DSG_Rejoj1
- ELV_Rejoj2
- DSG_Rejoj2
- ISG_Rejoj2

Ver Tabla 15. Modelo de Doble Interacción.

Coefficientes	Estimado	Error estándar	t-valor	Pr(> t)	
Intercepto	4,26E-08	2,87E-09	14,82	< 2,22E-16	***
MJDdecimal_Regresion	5,08E-10	1,19E-10	4,25	2,43E-05	***
REFGPS_Reloj1	4,94E-01	2,03E-02	24,30	< 2,22E-16	***
REFGPS_Reloj2	-5,22E-01	2,33E-02	-22,40	< 2,22E-16	***
ELV_Reloj1	-8,47E-12	4,52E-12	-1,87	0.0614	.
DSG_Reloj1	1,39E-11	7,63E-12	1,82	0.0688	.
ELV_Reloj2	5,63E-12	4,49E-12	1,25	0.2104	
DSG_Reloj2	5,32E-11	2,59E-11	2,04	0.0408	*
ISG_Reloj2	2,14E-11	7,74E-12	2,76	0.0058	**

Códigos de significatividad: 0 '***' 0,001 '**' 0,01 '*' 0,05 '.' 0,1 '' 1
 Error estándar de residuos: 7,99E-10 con 551 grados de libertad
 R-cuadrado múltiple: 0,5440; R-cuadrado ajustado: 0,5374
 Estadístico-F: 82,18 con 8 y 551 GL; p-valor: < 2,22E-16

Tabla 15. Modelo de Doble Interacción.

Este Modelo Doble Interacción logra explicar un 53,74% la variabilidad de los datos. Se trata de mejorar dicho porcentaje y se recurre a la técnica de "stepwise" y limitando las interacciones múltiples solamente a estas variables.

Coefficientes	Estimado	Error estándar	t-valor	Pr(> t)	
Intercepto	2,06E-7	5,88E-08	3,49	0,0005	***
MJDdecimal_Regresion	-4,61E-10	3,69E-10	-1,24	0,2118	
REFGPS_Reloj1	2,15	7,05E-01	3,04	0,0024	**
REFGPS_Reloj2	5,05E-01	3,46E-01	1,45	0,1454	
ELV_Reloj1	-9,81E-11	2,54E-11	-3,86	0,0001	***
DSG_Reloj1	-1,66E-11	1,36E-11	-1,21	0,2236	
ELV_Reloj2	4,17E-11	1,07E-11	3,91	0,0001	***
DSG_Reloj2	2,86E-10	7,87E-11	3,63	0,0003	***
ISG_Reloj2	-1,02E-10	2,02E-10	-0,50	0,6146	
REFGPS_Reloj1:ELV_Reloj2	4,42E-04	1,24E-04	3,57	0,0003	***
REFGPS_Reloj2:ELV_Reloj1	-5,54E-04	1,46E-04	-3,79	0,0001	***
MJDdecimal_Regresion:DSG_Reloj1	5,45E-11	2,06E-11	2,64	0,0084	**
REFGPS_Reloj1:REFGPS_Reloj2	1,04E+07	4,12E+06	2,51	0,0121	*
REFGPS_Reloj1:ISG_Reloj2	-3,70E-03	2,47E-03	-1,49	0,1351	
DSG_Reloj2:ISG_Reloj2	-1,20E-11	3,64E-12	-3,29	0,0010	**
ELV_Reloj1:ISG_Reloj2	-1,96E-13	6,51E-14	-3,01	0,0027	**
ELV_Reloj1:ELV_Reloj2	-4,04E-15	1,73E-15	-2,33	0,0201	*

Códigos de significatividad: 0 '***' 0,001 '**' 0,01 '*' 0,05 '.' 0,1 '' 1
 Error estándar de residuos: 7,71E-10 con 543 grados de libertad
 R-cuadrado múltiple: 0,5814; R-cuadrado ajustado: 0,5691
 Estadístico-F: 47,15 con 16 y 543 GL; p-valor: < 2,22E-16

Tabla 16. Modelo de Regresión mediante la técnica "stepwise"

El modelo se vuelve más complejo y solo se consigue un 3 % de mejora respecto del nuevo modelo base.

Por lo anteriormente explicado, se elige el Modelo Doble Interacción. El mismo tiene la siguiente fórmula:

$$Prom_{DifREFGPS} = \beta_0 + \beta_1 MJDdecimal_{Regresion} + \beta_2 S_{Reloj1} + \beta_3 REFGPS_{Reloj2} + \beta_4 ELV_{Reloj1} + \beta_5 DSG_{Reloj1} + \beta_6 ELV_{Reloj2} + \beta_7 DSG_{Reloj2} + \beta_8 ISG_{Reloj2} + \varepsilon$$

COMPROBACIÓN DE LOS SUPUESTOS DEL MODELO

A diferencia de una técnica exploratoria, un Modelo de Regresión Múltiple tiene supuestos, los cuales deben ser verificados. Los supuestos del modelo son:

1. *Linealidad*
2. $E(\varepsilon) = 0$
3. $Var(\varepsilon) = \sigma^2 I_n$
4. *Los errores no están correlacionados*
5. *El vector ε se distribuye $Normal_{(0; \sigma^2 I_n)}$*

Aclaración la variable ε hace referencia a los residuos de la regresión lineal múltiple

Para analizar los residuos conviene observar los residuos escalados, en este caso se observan los residuos estudentizados, los mismos deben estar acotados entre -2 y 2.

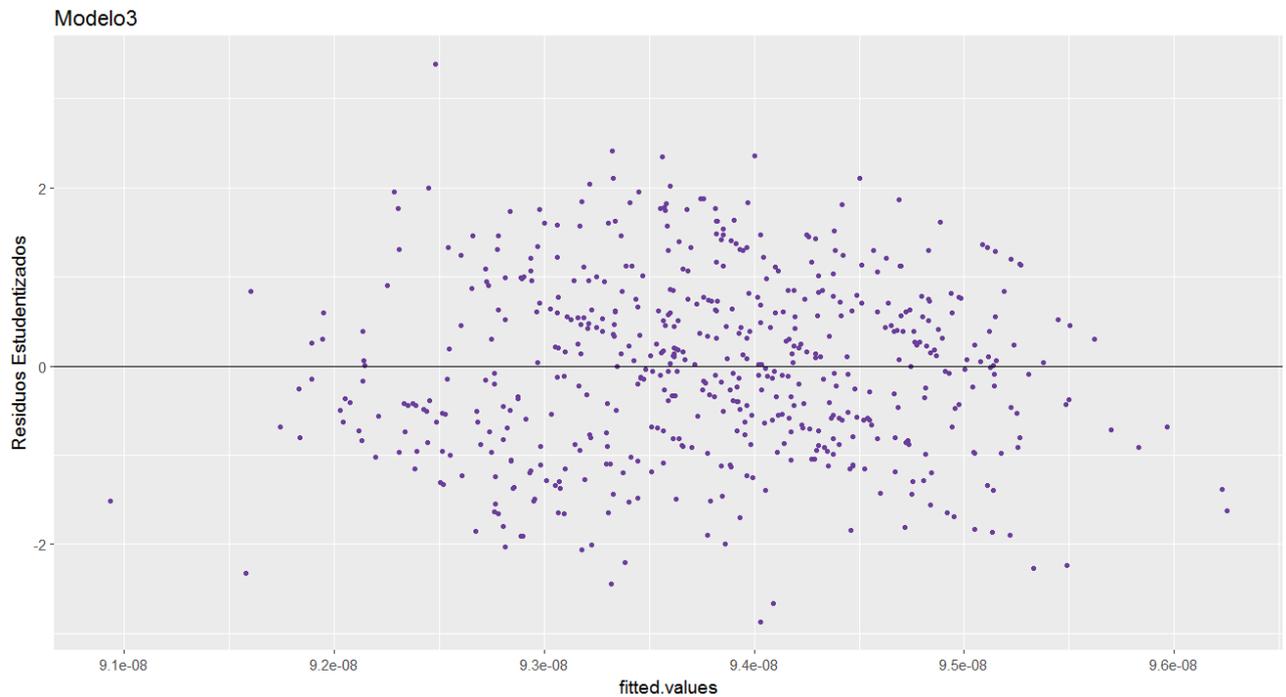


Figura 37. Valores ajustados vs. Residuos estudentizados del Modelo Doble Interacción

Se observa una nube de puntos sin estructura, lo cual indica que no hay correlación alguna. La nube está centrada en torno a los residuos estudentizados con valor cercano a cero y contiene a la mayoría de los puntos entre -2 y 2. Esto nos da una idea que la varianza es constante. Existen algunos puntos fuera de estos límites pero no influyen de manera significativa.

Con los mismos criterios, se puede observar los residuos estudentizados contra las variables predictoras.

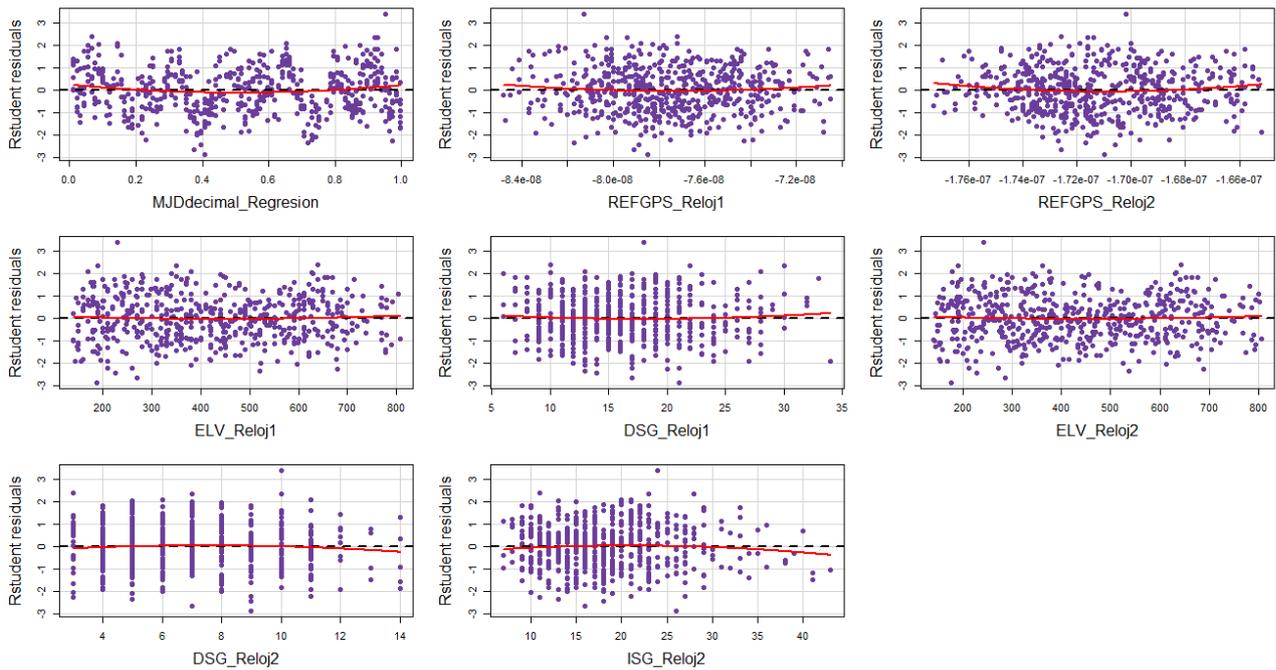


Figura 38. Variables predictorias vs. Residuos estudentizados del Modelo Doble Interacción

Para analizar normalidad de los residuos, se puede hacer mediante, un diagrama de caja, un histograma o un gráfico de normalidad Q-Q.

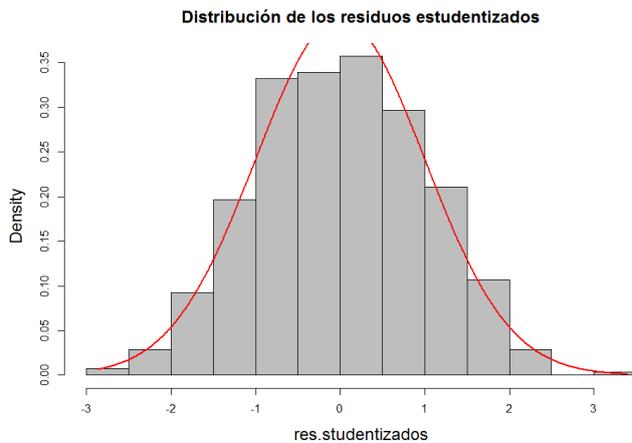


Figura 39. Histograma de los residuos estudentizados del Modelo Doble Interacción

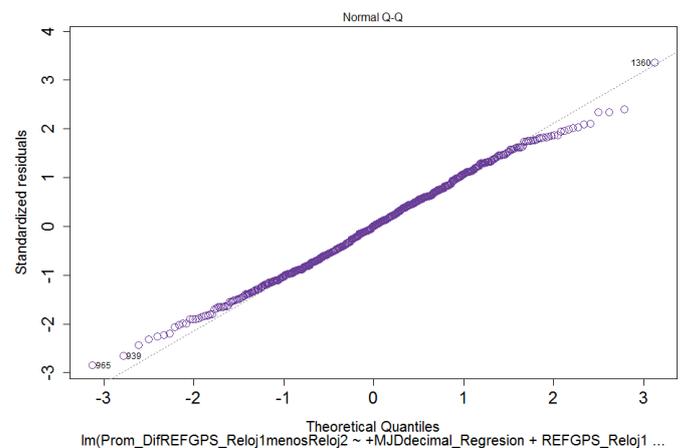


Figura 40. Gráfico Q-Q de Histograma de los residuos estudentizados del Modelo Doble Interacción

En este caso se opta por un histograma de los residuos estudentizados, al cual se le agregó una distribución normal como referencia - color rojo -.

Para complementar el análisis, se optó por un gráfico de Q-Q, Figura 40. Este tipo de gráfico, nos indica que si los puntos están sobre la diagonal, se tiene una distribución normal. Se observa que la mayoría de los puntos se ubica sobre dicha diagonal, sin embargo algunos puntos se alejan de dicha recta.

Luego de realizar una regresión, incluso habiendo hecho una limpieza de datos, pueden existir algunos registros que afecten la forma en la que realiza el ajuste lineal, o pueden hacer que no se cumplan los supuestos. Dichos registros son conocidos como puntos influyentes.

Una forma de cuantificar y encontrar puntos influyentes es mediante la distancia de Cook. La misma mide el cambio que ocurriría en el vector de coeficientes estimados de regresión si la i -ésima observación fuese omitida. Se calcula como:

$$D_i = \frac{\sum_{j=1}^n (\hat{y}_j - \hat{y}_{j(i)})^2}{\hat{\sigma}^2(p+1)} \quad i = 1, 2, \dots, n$$

con p : variables regresoras

Aquellas observaciones cuya distancia de Cook exceda determinado valor pueden ser consideradas influyentes. Existen diversos criterios para determinar un punto de corte o umbral, algunos sugieren que una observación sea considerada como influyente si es mayor a $4/n$, ó a 1. En la Figura 41 se puede observar un punto de corte, color rojo, cuyo valor está dado por:

$$\text{Punto de corte} = \frac{4}{\text{nro. de registros}}$$

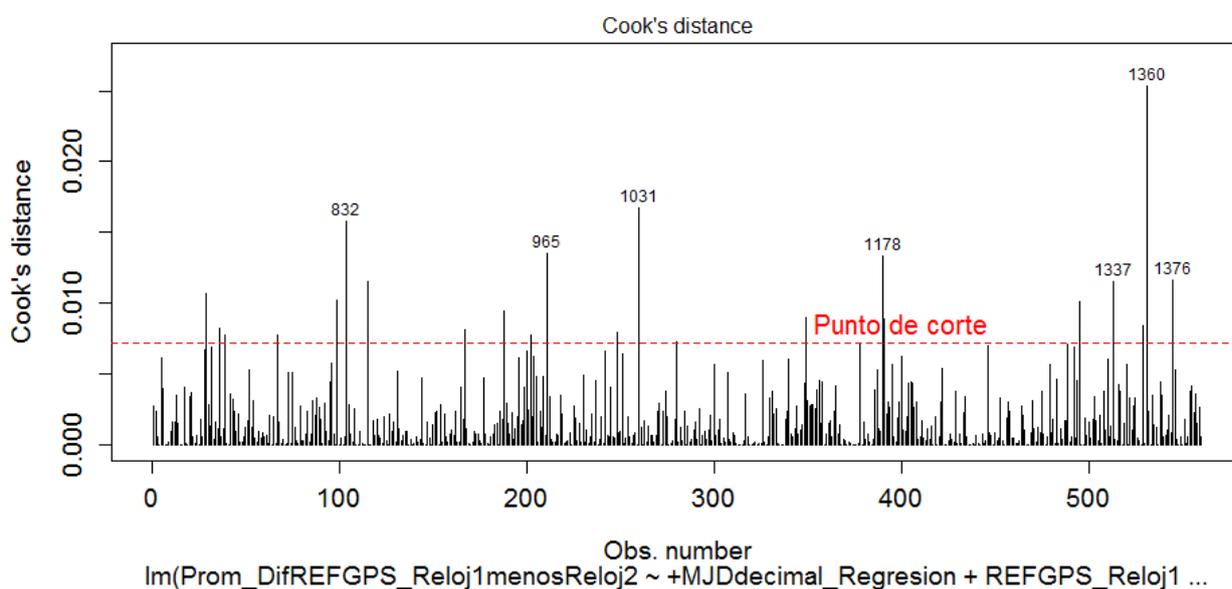


Figura 41. Distancias de Cook para el Modelo Doble Interacción

Si bien la Figura 41 el punto de corte nos indica varios puntos influyentes, en general, se recomienda analizar aquellos puntos de mayor valor que queden alejados del resto [7]. Si se ordena en forma decreciente el vector de distancias de Cook, se observa que la mayor distancia de Cook es alcanzada por la observación 1360, dicho valor se reduce drásticamente a la mitad para la observación 1376. En base a este criterio se identificaron y analizaron los primeros siete registros con mayor distancia de Cook.

Todo lo anteriormente dicho permite validar los supuestos del modelo. Sin embargo, es importante estudiar la multicolinealidad en el conjunto de datos. Algunos efectos que puede causar la multicolinealidad son:

- Los estimadores de los coeficientes tienen grandes desvíos, lo cual afecta la inferencia y predicción.
- Es poco acertado dar la interpretación usual a los coeficientes de la regresión ajustada (por ejemplo dan con signo contrario al esperado).
- Los valores estimados para los coeficientes son sensibles a pequeños cambios en los datos o al hecho de quitar/agregar alguna variable.
- Los coeficientes estimados pueden dar muy grandes en valor absoluto.

Una herramienta para estudiar la multicolinealidad, es el mediante un correlograma. Ver Figura 42. Correlograma de las variables del Modelo Doble Interacción

Correlograma de las variables del Modelo

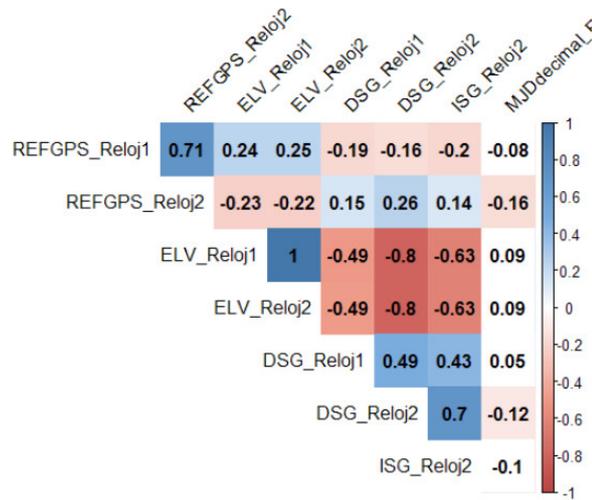


Figura 42. Correlograma de las variables del Modelo Doble Interacción

Las variables de elevación del satélite, ELV_Relej1 y ELV_Relej2, tienen una correlación muy alta. Esto ya implica tener especial cuidado con dichas variables.

El problema del correlograma es que estudia las variables tomadas de a dos. Una herramienta alternativa puede ser el estudio de VIF - Factor de Inflación de la Varianza -. El VIF para la variable regresora X_j se calcula del siguiente modo:

$$VIF_j = \frac{1}{1 - R_j^2} \quad j = 1, 2, \dots, p.$$

Donde R_j^2 es el coeficiente de determinación de X_j versus las restantes regresoras.

Si el VIF_j es igual a uno, entonces el $R_j^2 = 0$, por lo tanto la variable X_j no está correlacionada con las restantes covariables. Cuando el $R_j^2 \neq 0$, el VIF_j es mayor a uno. Cuando R_j^2 está muy cerca de uno, el VIF_j tiende a infinito, razón por la cual algunas veces se utiliza la inversa del VIF_j :

$$\text{Índice de tolerancia} = \frac{1}{VIF_j}$$

Otra herramienta para analizar multicolinealidad, es definir el Número de Condición a partir de la matriz de correlación de $X^T * X$. Dado que se trata de una matriz cuadrada, se puede definir el número de condición en base a sus autovalores:

$$\text{Número de Condición} \rightarrow K = \sqrt{\frac{\lambda_{MAX}}{\lambda_{min}}}$$

El Número de Condición siempre será mayor a 1. Si el Número de Condición excede a valores mayores a 15, existe un indicio de multicolinealidad y los datos deben ser analizados en detalle⁴. Acciones correctivas deben tomarse cuando el número de condición de la matriz de correlación excede valores mayores a 30.

⁴ Aclaración: El valor de corte de 15 no se basa en ninguna consideración teórica, sino que surge de las observaciones empíricas [6].

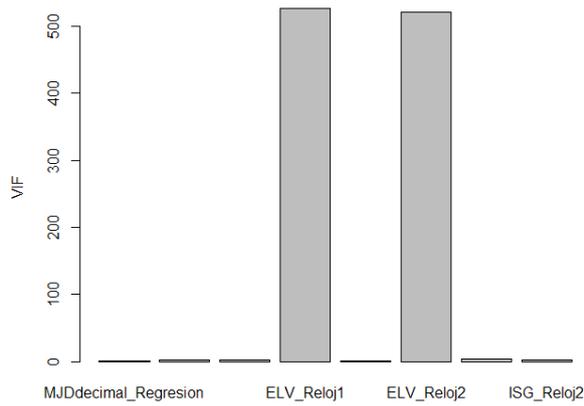


Figura 43. Valores de VIF variables del Modelo Doble Interacción

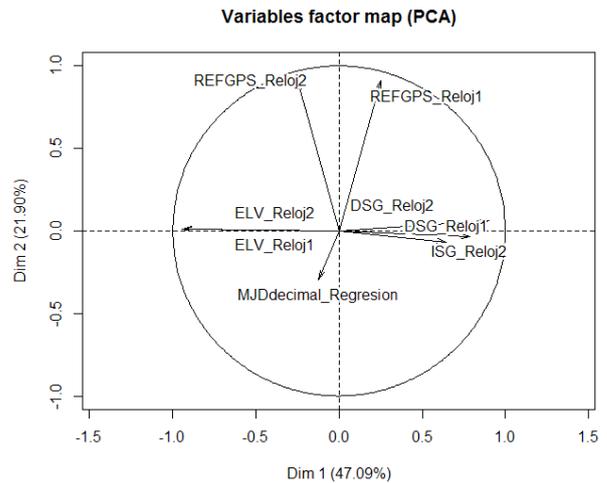


Figura 44. PCA sobre variables del Modelo Doble Interacción

Si los valores de VIF son mayores a 10, es muy probable que hay multicolinealidad entre dichas variables.

Si el número de condición es mayor a 100 se presenta una multicolinealidad moderada y con valores mayores a 1000 una multicolinealidad grave.

Se puede apreciar que el valor de VIF de las variables de elevación tiene un valor mayor a 500. Por otro lado si se aplica la técnica de Componentes Principales y se observa un gráfico x-y con sus dos componentes mayoritarias, es fácil de apreciar que los vectores de dichas variables tienen el mismo sentido y misma dirección. Por otro lado el número de condición es de 3938,87.

A continuación se detallan los beta estandarizados para el Modelo Doble Interacción.

Betas estandarizados							
Intercepto	MJDdecimal_Regresion	REFGPS_Relej1	REFGPS_Relej2	ELV_Relej1	DSG_Relej1	ELV_Relej2	DSG_Relej2
0,000	0,126	1,252	-1,160	-1,237	0,064	0,823	0,112

Tabla 17. Betas estandarizados del Modelo Doble Interacción

Se debe apreciar que los valores no son muy elevados, pero sin embargo llama la atención el signo contrario que tienen las variables ELV_Relej1 y ELV_Relej2.

Si se vuelve a pensar en el del Modelo Doble Interacción, esta vinculación entre ambas variables es lógica. Esto posiblemente se deba a que las antenas GPS están muy cercanas entre sí, y para Vista Común ambos Sistemas de Medición deben visualizar al mismo satélite. Todo lo anteriormente expuesto, se considera suficiente evidencia para descartar este modelo.

UN MODELO POR CADA SISTEMA DE MEDICIÓN

SISTEMA DE MEDICIÓN 1

Ya se sabe que las interacciones tomadas de dos, mejoran el modelo pero resulta complicado darle una explicación a dicha variable. Se opta por el método de simplificación de variables “backward”, partiendo de un modelo que contiene todas las variables del Sistema de Medición Nro. 1. Finalmente se llega al siguiente modelo:

Coefficientes	Estimado	Error Estándar	t-valor	Pr(> t)	
Intercepto	-4,82E-08	1,64E-09	-29,34	< 2,22E-16	***
REFGPS_Reloj1	3,54E-01	1,99E-02	17,78	< 2,22E-16	***
ELV_Reloj1	-2,62E-12	7,26E-13	-3,60	0,0003	***
DSG_Reloj1	-3,57E-11	1,18E-11	-3,01	0,0027	**
IOE_Reloj1	-6,27E-12	1,80E-12	-3,48	0,0005	***
MDTR_Reloj1	7,28E-02	2,04E-02	3,57	0,0003	***
MDIO_Reloj1	-1,07E-01	1,40E-02	-7,65	8,84E-14	***
MJDdecimal_Regresion	1,01E-09	2,51E-10	4,00	6,91E-05	***

Códigos de significatividad: 0 '***' 0,001 '**' 0,01 '*' 0,05 '.' 0,1 ' ' 1
 Error estándar de residuos: 1,27E-09 con 552 grados de libertad
 R-cuadrado múltiple: 0,5155, R-cuadrado ajustado: 0,5093
 Estadístico-F: 83,91 con 7 y 552 GL, p-valor: < 2.22E-16

Tabla 18. Modelo de Regresión para el del Sist. de Med. Nro. 1

Si bien la variabilidad no es muy elevada 50,93%, es de notar que todas las variables son significativas.

COMPROBACIÓN DE LOS SUPUESTOS SISTEMA DE MEDICIÓN NRO 1

Se observa la distribución de los residuos estudentizados.

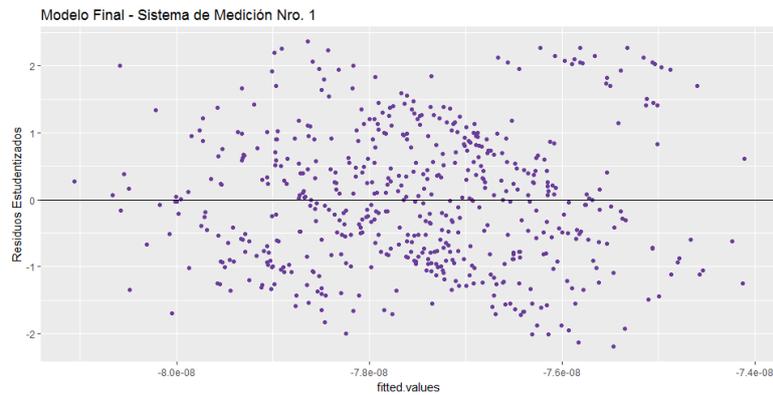


Figura 45. Residuos estudentizados vs. Valores ajustados para Modelo del Sistema de Medición 1

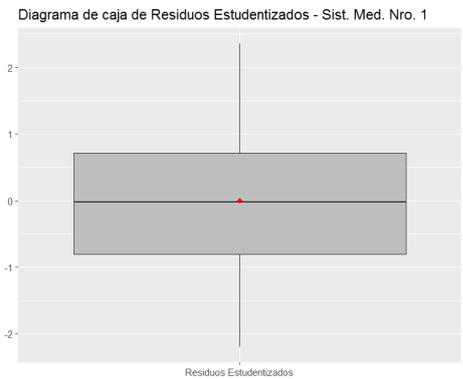


Figura 46. Diagrama de caja de los residuos estudentizados para Modelo del Sistema de Medición 1

Residuos Estudentizados. Estadísticos descriptivos					
Mínimo	Primer cuartil	Mediana	Media	Tercer cuartil	Máximo
-2,20	-0,80	-0,03	0,00016	0,71	2,36

Tabla 19. Estadísticos descriptivos de los residuos estudentizados del Modelo del Sist. de Med. 1

En la Figura 45 caso se graficaron los residuos estudentizados vs. los valores ajustados, también se utilizó un diagrama de caja y finalmente se en la Tabla 19 se muestran los estadísticos descriptivos. Mediante los datos presentados se concluye que los residuos están acotados, su varianza es constante y conforman una nube de puntos sin forma alguna.

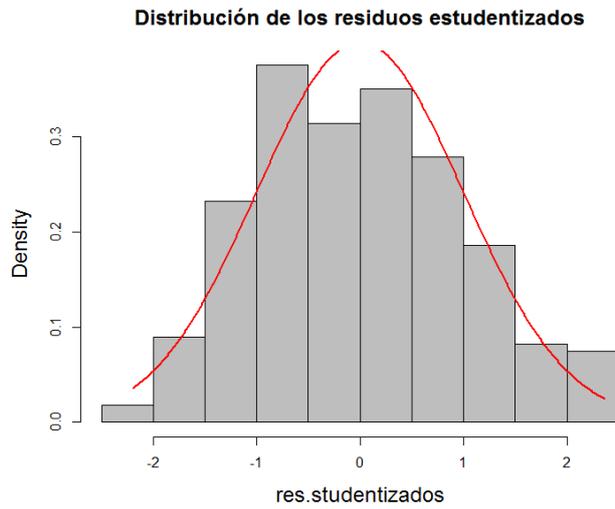


Figura 47. Histograma de los residuos estudentizados del Modelo del Sistema de Medición Nro. 1

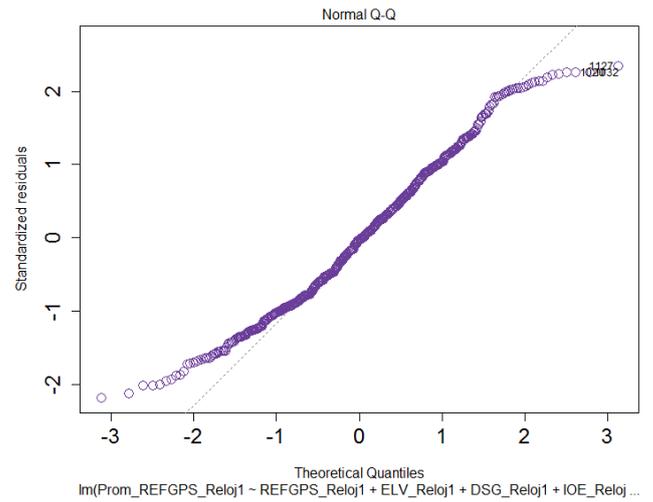


Figura 48. Gráfico Q-Q residuos estudentizados del Modelo del Sistema de Medición Nro. 1

La normalidad de los residuos es cuestionable, ya que el intervalo de puntos que caen sobre la diagonal del gráfico Q-Q es muy chico.

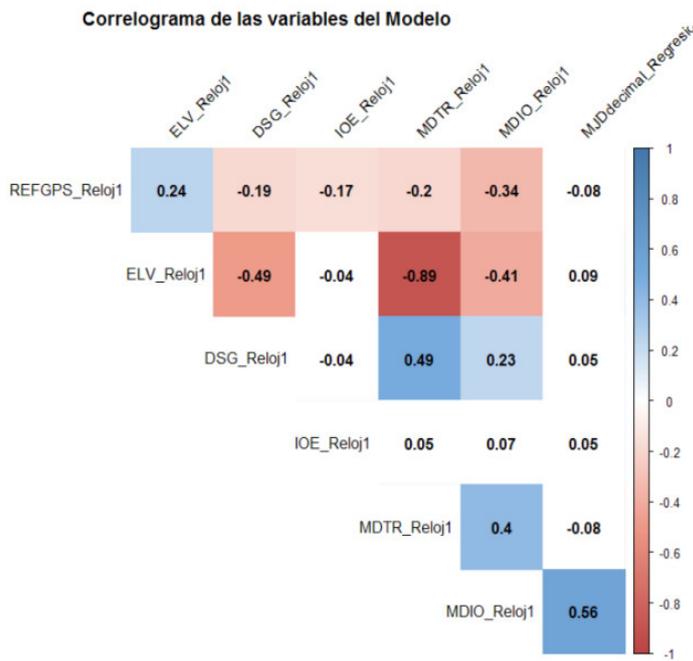


Figura 49. Correlograma de las variables Modelo del Sistema de Medición Nro. 1

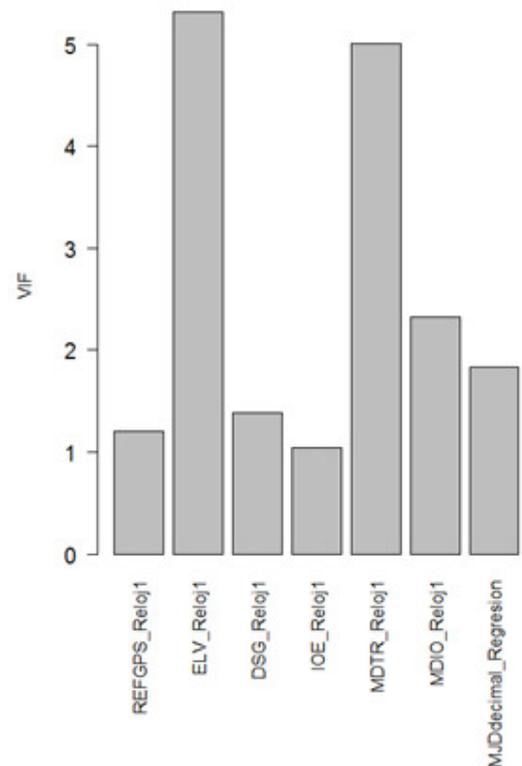


Figura 50. VIF de las variables Modelo del Sistema de Medición Nro. 1

La única correlación elevada es la que vincula a ELV_Reloj1 con MDTR_Reloj1 y dicho esto ambas variables presentar un valor de VIF cercano a 5. Si se elimina una de estas dos variables, se llega a una instancia en la que ciertas variables se vuelven poco significativas, por ejemplo: MJDdecimal_Regresion la cual implica el paso del tiempo a lo largo del día. Es por esto que, a pesar que el modelo no cumple con todos los supuestos en su justa medida, se acepta este modelo como válido.

SISTEMA DE MEDICIÓN NRO. 2

Bajo el mismo criterio, método "backward" partiendo de todas las variables, se genera el modelo para el sistema de medición.

Coefficientes	Estimado	Error estándar	t-valor	Pr(> t)	
Intercepto	-1,13E-07	3,39E-09	-33,49	< 2,22E-16	***
REFGPS_Relej2	3,33E-01	1,96E-02	16,96	< 2,22E-16	***
MDTR_Relej2	8,31E-02	1,21E-02	6,88	1,60E-11	***
SMDT_Relej2	-1,02E+02	2,40E+01	-4,24	2,60E-05	***
SMDI_Relej2	2,52E+02	5,04E+01	5,00	7,71E-07	***
MDIO_Relej2	-1,48E-01	1,23E-02	-12,00	< 2,22E-16	***
ISG_Relej2	-3,53E-11	9,55E-12	-3,69	0,0002	***
MJDdecimal_Regresion	9,49E-10	2,24E-10	4,22	2,76E-05	***

Códigos de significatividad: 0 '***' 0,001 '**' 0,01 '*' 0,05 '.' 0,1 ' ' 1
 Error estándar de residuos: 1,10E-09 con 552 grados de libertad
 R-cuadrado múltiple: 0,5704; R-cuadrado ajustado: 0.5650
 Estadístico-F: 104,73 con 7 y 552 GL, p-valor: < 2,22E-16

Tabla 20. Modelo de Regresión para el del Sist. de Med. Nro. 2

Aquí todas las variables tienen un p -valor < 0.001, sin embargo para llegar a este tipo de modelo se tuvo que adoptar otras variables que no coinciden con las variables adoptadas en el Sistema de Medición Nro. 1.

Esto dificulta la comprensión de los modelos, ya que se trata de sistemas que miden la misma magnitud, con la misma tecnología, mismo principio de funcionamiento, en las mismas condiciones ambientales y prácticamente sus antenas no distan más de dos metros de distancia.

COMPROBACION DE LOS SUPUESTOS SISTEMA DE MEDICIÓN NRO. 2

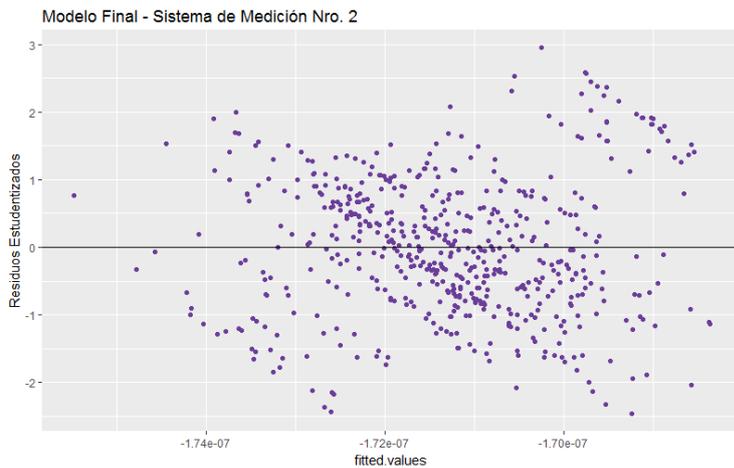


Figura 51. Residuos estudentizados vs. Valores ajustados para Modelo del Sistema de Medición Nro. 2

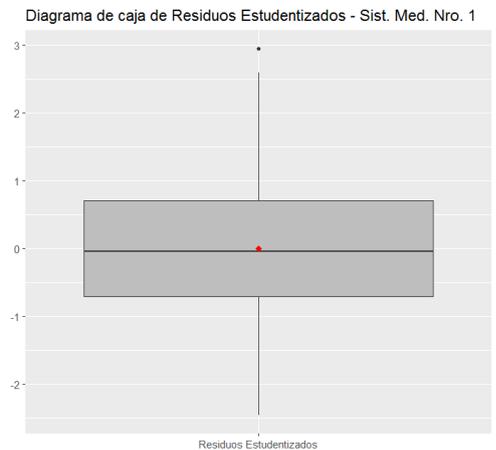


Figura 52. Diag. de caja de los residuos estudentizados para Modelo del Sist. de Medición Nro. 2

En la Figura 51, se observa que la nube de puntos presenta homogeneidad en su varianza, ya que la misma no crece ni disminuye. El diagrama de caja, nos indica que su media esta levemente corrida del cero.

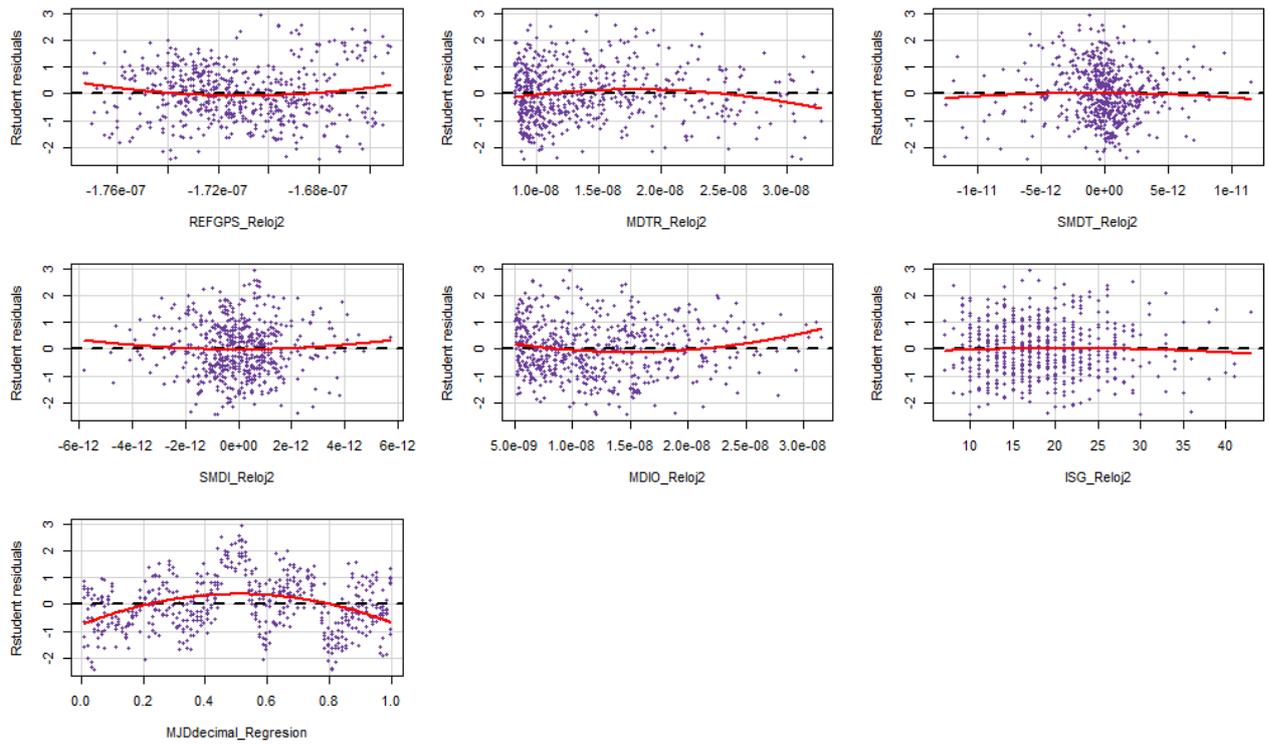


Figura 53. Residuos estudentizados vs. variables predictoras del Modelo del Sistema de Medición Nro. 2

El gráfico de los residuos estudentizados vs. las variables predichas tiene algunos ajustes de tipo parabólico pero parecen compensarse entre sí.

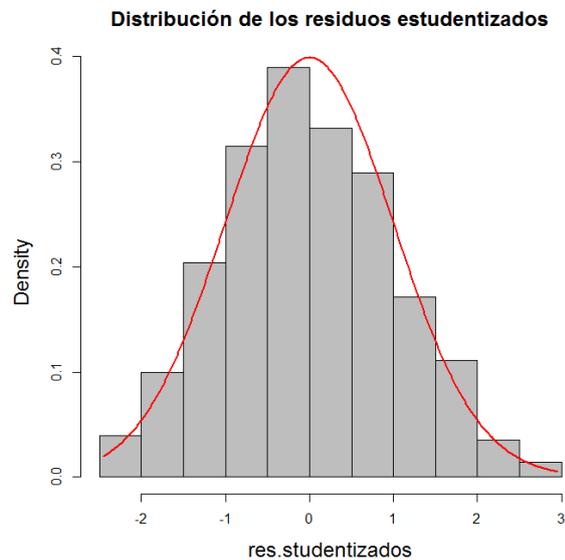


Figura 54. Histograma de los residuos estudentizados del Modelo del Sistema de Medición Nro. 2

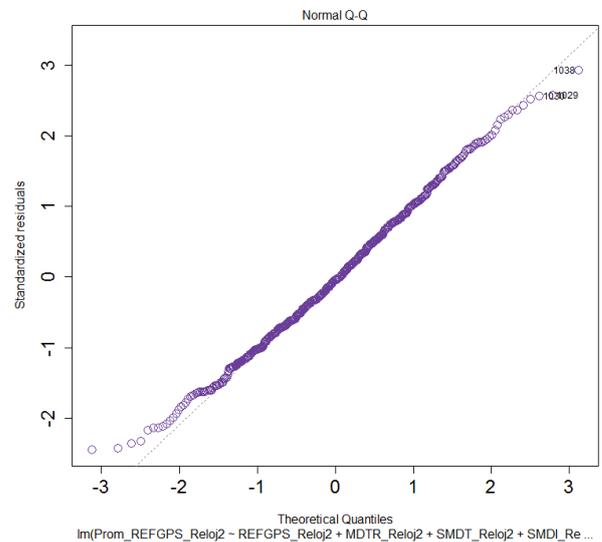


Figura 55. Gráfico Q-Q residuos estudentizados del Modelo del Sistema de Medición Nro. 2

En el histograma la normalidad de los residuos se puede ver que está comprometida en su cola izquierda. Eso se evidencia en el intervalo donde del gráfico Q-Q donde el intervalo de puntos que se posiciona sobre la diagonal va desde -1 hasta +2.

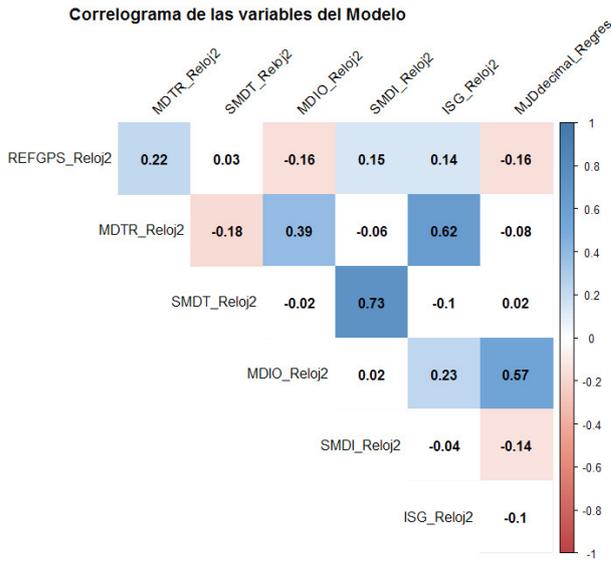


Figura 56. Correlograma de las variables Modelo del Sistema de Medición Nro. 2

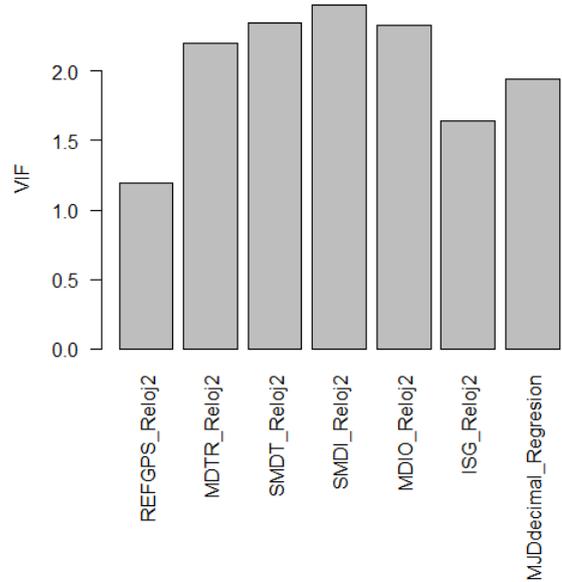


Figura 57. VIF de las variables Modelo del Sistema de Medición Nro. 2

Respecto de la multicolinealidad, se puede apreciar que ninguna supera el módulo de 0,8; y los valores de VIF son inferiores a 3. No hay sospecha alguna de multicolinealidad.

Aunque las variables predictoras de este modelo difieran modelo elegido para el Sistema de Medición Nro. 1, se adopta el modelo recientemente estudiado para el Sistema de Medición Nro. 2.

APLICACIÓN DE AMBOS MODELOS

Se debe comprender que la variable predicha de cualquiera de los modelos, no es la variable de interés. La variable a estimar, surge de la diferencia de los valores ajustados de los modelos elegidos.

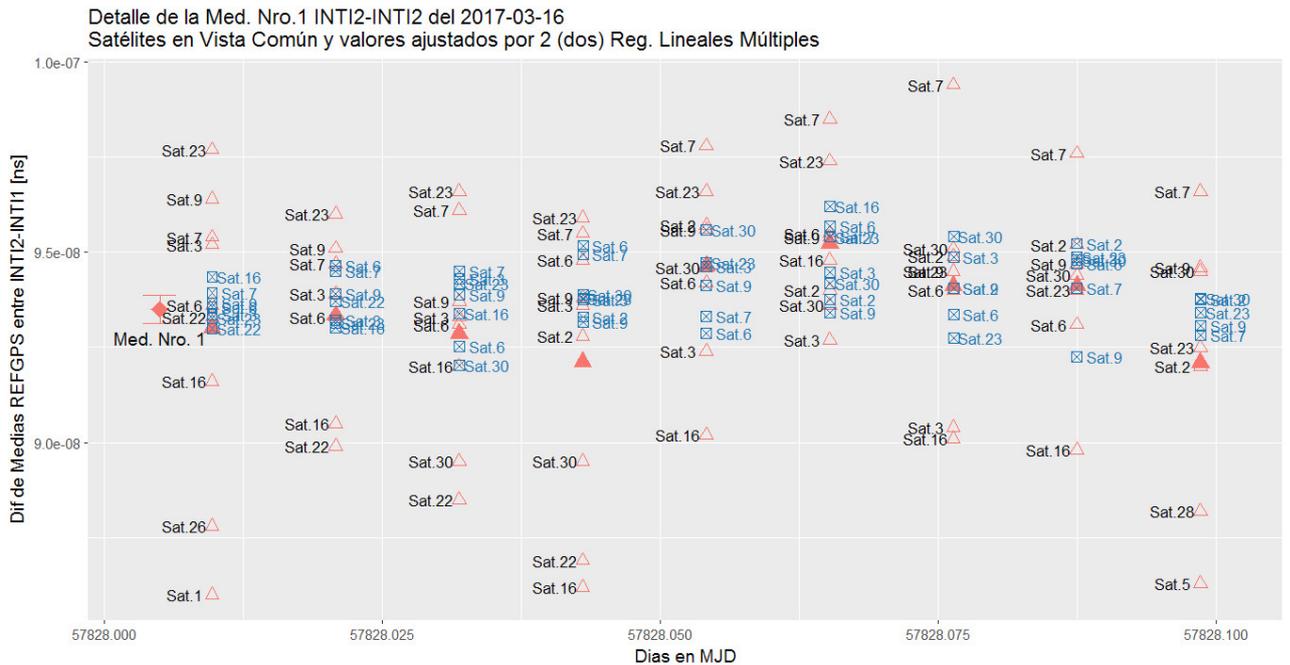


Figura 58. Med. Nro 1 del 2017-03-16. Valores promedio de VC 16 min. y diferencia de valores ajustados.

En la Figura 58 se observa en color azul, la diferencia de los valores ajustados para cada modelo. Es fácil de apreciar que los nuevos valores están menos dispersos.

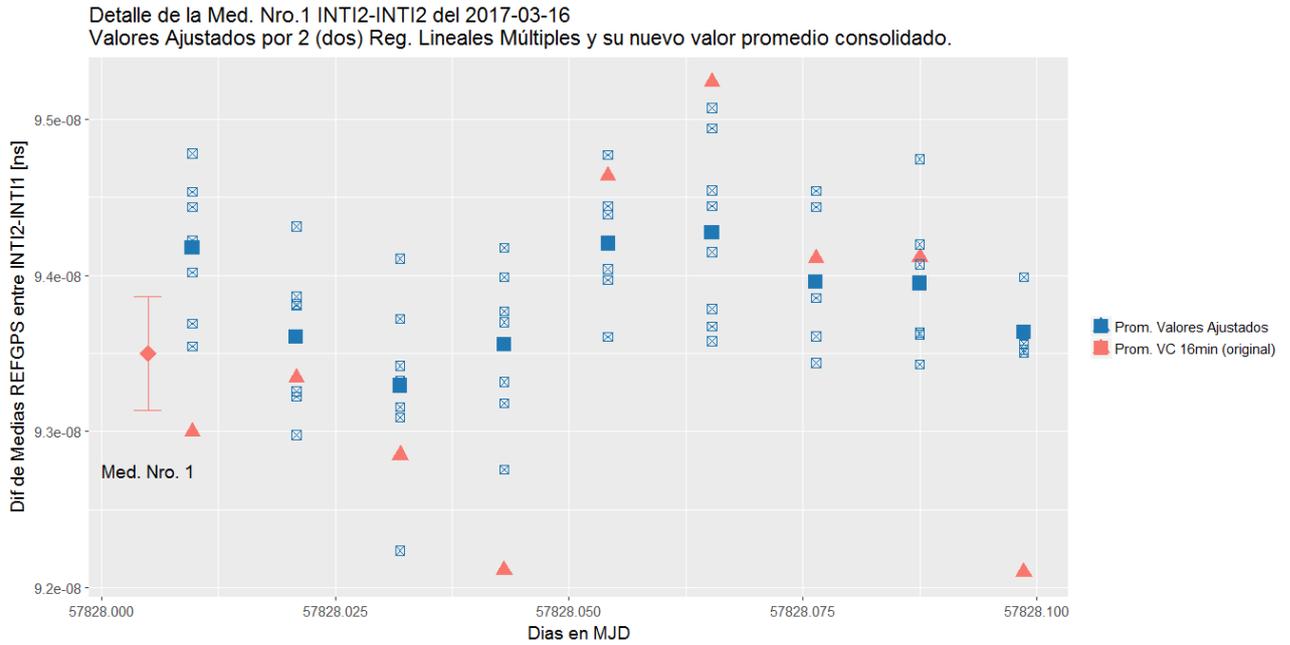


Figura 59. Med. Nro 1 del 2017-03-16. Promedios de VC 16 min. y nuevo promedio de valores ajustados.

Tal como se anticipó en el análisis de regresión, se puede observar que para la Figura 59 el promedio de la diferencia de los valores ajustados - color azul, marcador cuadrado lleno-, no coincide con los valores promedios de Vista Común cada 16 minutos - color rojo, marcador triángulo lleno -.

Este nuevo valor de medida, cuadrado azul observado en la Figura 59, resulta de promediar los valores ajustados por el modelo de regresión lineal múltiple. Otra alternativa al promedio de los valores ajustados por la regresión, es utilizar:

- Mediana de los valores ajustados
- Promedio ponderando la inversa de los residuos de la regresión
- Promedio ponderando la inversa de la diferencia de la variable DSG para ambos sistemas.

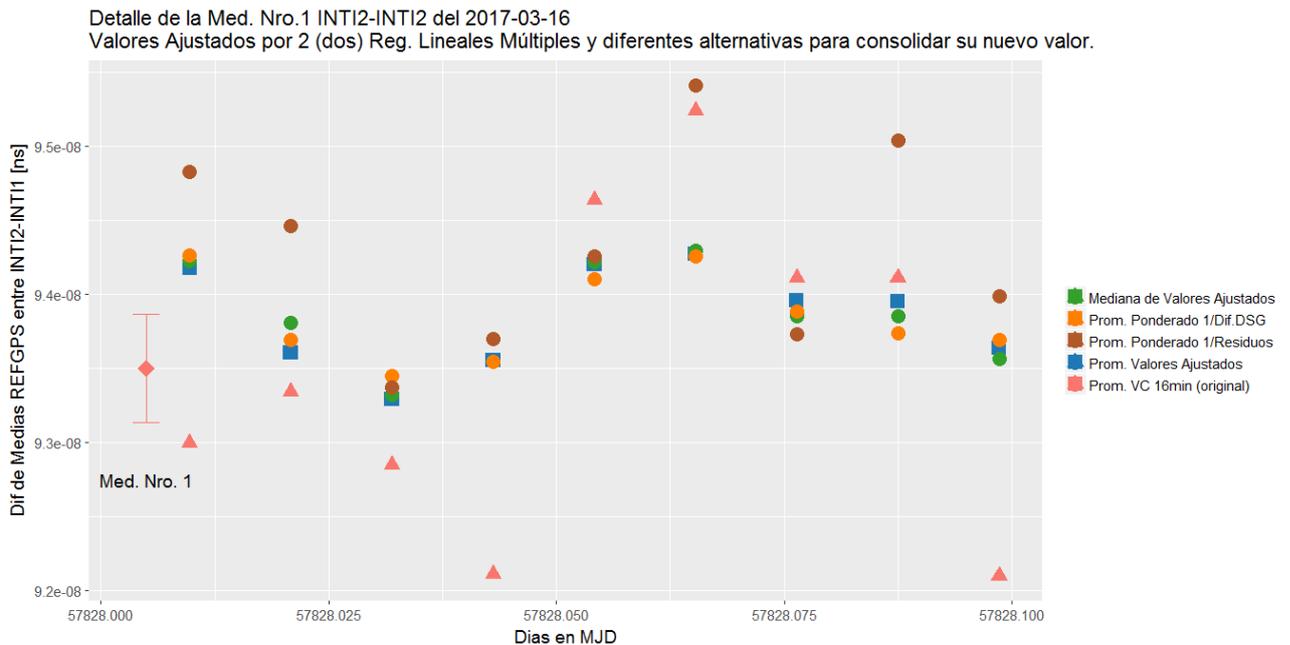


Figura 60. Med. Nro 1 del 2017-03-16. Promedios de VC 16 min. y alternativas para consolidar su nuevo valor.

RESULTADOS

Para ponderar los resultados alcanzados se recurre al gráfico de control y el Desvío de Allan.

Comparación de los relojes INTI2 e INTI1 desde el 16 Marzo 2017

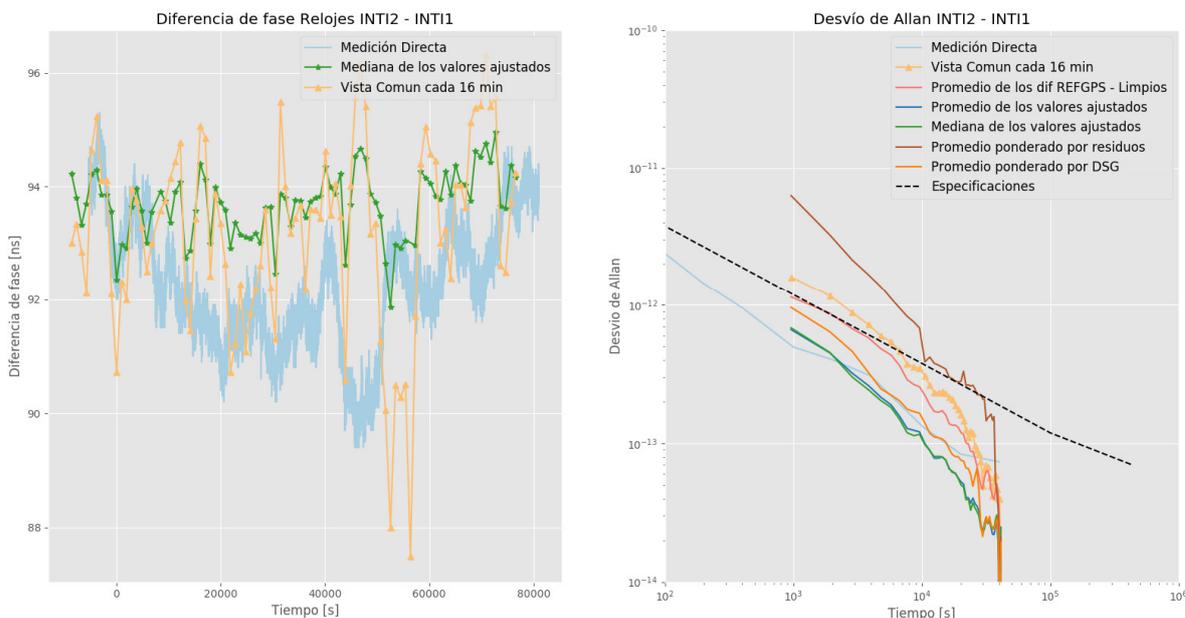


Figura 61. Resultados. A izq.: Comparación en el tiempo. A derecha: Desvío de Allan

En el gráfico de control de la Figura 61, se puede observar que mediante la regresión lineal múltiple para ambos sistemas y consolidando los resultados mediante la mediana - color verde - se logra generar una curva se puntos más suave. El Desvío de Allan para 16 minutos cumple las especificaciones tanto para la mediana como para el promedio. Es una decisión de este trabajo optar y sugerir el uso de la mediana. Se puede observar que la limpieza de datos -color rojo - mejora la estabilidad de nuestro sistema. Ponderar mediante los residuos no ha resultado efectivo.

El modelo fue entrenado con los datos del día 16 de Marzo de 2017, a continuación se evalúan los resultados en función de la capacidad predictiva del modelo para el día 18 de Marzo 2017.

Predicción de los relojes INTI2 e INTI1 para el 18 Marzo 2017

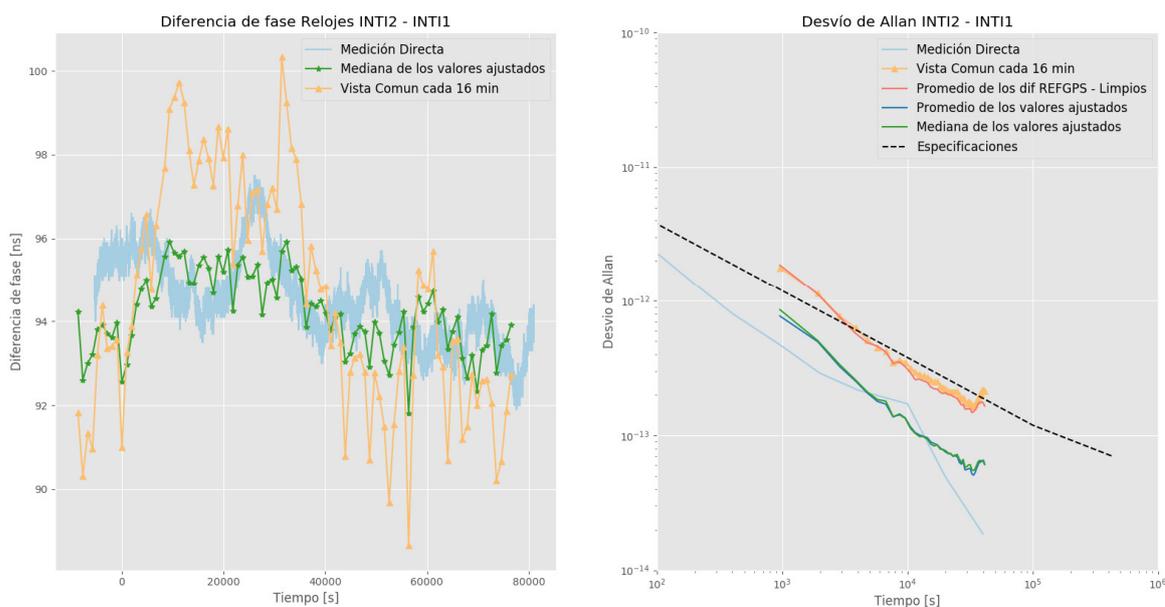


Figura 62. Resultados de la predicción

Se observa la que la capacidad predictiva del modelo funciona correctamente - no se presentan mayores diferencias en el gráfico de control y la Desvío de Allan cumple con las especificaciones requeridas.

CONCLUSIONES

Se pudo demostrar que la limpieza de datos es necesaria y permite detectar datos anómalos presentes en el conjuntos de datos. La regresión tiene por objetivo disminuir la distancia de los residuos, esto hace que la dispersión existente entre los satélites disminuya. Finalmente el reemplazo del promedio por la mediana hace que consigan mediciones más estables para Vista Común cada 16 minutos. Esto es disminuye el tiempo de la ventana de integración aproximadamente 10 veces, respecto de Vista Común cada 144 minutos.

Si bien es cierto que la capacidad predictiva del modelo ha funcionado correctamente, se debe tener en cuenta que no se pudo explicar por qué los modelos generados, en base a los datos, para cada Sistema de Medición fue distinto. Esto es algo que debe seguir siendo objeto de estudio. En base a una mejor comprensión del modelo se pueden desarrollar mejores herramientas y criterios para la limpieza de datos.

AGRADECIMIENTOS

Cabe resaltar la buena predisposición del Laboratorio de Tiempo y Frecuencia del Instituto Nacional de Tecnología Industrial, quienes compartieron los datos de manera abierta y se ofrecieron a realizar técnica de Medición Directa durante el período de estudio.

En especial al Dr. Diego Luna, quien cumplió un vital aporte de documentación de equipos, artículos, publicaciones de interés, recomendaciones, puntos de vista y consejos.

¡Muchas gracias!

REFERENCIAS

- [1] T. Gotoh, A. Kaneko, Y. Shibuya, and M. Imae, "GPS common view," *J. Nat. Inst. Inf. Commun. Technol.*, vol. 50, nos. 1–2, pp. 113–123, 2003.
- [2] Luna, D., Pérez, D., Cifuentes, A., & Gómez, D. (2017). Three-Cornered Hat Method via GPS Common-View Comparisons. *IEEE Transactions on Instrumentation and Measurement*.
- [3] Ana M. Bianco y Elena J. Martínez. Apunte de clase. Probabilidades y Estadística (Computación). Facultad de Ciencias Exactas y Naturales. Universidad de Buenos Aires.
- [4] Costa, R., Orgiazzi, D., Pettiti, V., Sesia, I., & Tavella, P. (2004). Performance comparison and stability characterization of timing and geodetic GPS receivers at IEN.
- [5] Luna, D. A., Pérez, D. N., Palacios, F., & Cifuentes, A. (2016, July). ANÁLISIS DE ESTABILIDAD EN RELOJES ATÓMICOS POR MEDIO DE LA TÉCNICA VISTA COMÚN GPS. In ANALES AFA (Vol. 27, No. 2).
- [6] Chatterjee, S., & Hadi, A. S. (2015). *Regression analysis by example*. John Wiley & Sons.
- [7] Norman R. Draper, Harry Smith. *Applied Regression Analysis*, 3rd Edition. Wiley, p. 212.