

Análisis de los datos de turbidez del ECO FLNTU correspondientes a la campañas de diciembre de 2019

Ramiro Santamaria, Santiago Scheiner

2020-06-01

Resumen

En este informe concentramos los diferentes análisis que hicimos para los datos de turbidez del ECO FLNTU correspondientes a la campañas de diciembre de 2019 en el Muelle de Pescadores de Palermo.

Comenzamos mostrando brevemente los resultados de la campaña del 2019-12-10, en la que tuvimos inconvenientes con los instrumentos, y mostramos por qué consideramos que los datos no eran suficientes para caracterizar la adquisición de datos del ECO FLNTU.

A continuación, mostramos varios pasos del proceso que fuimos realizando con los datos del instrumento a caracterizar a partir de los datos obtenidos en la campaña del 2019-12-17, detallando las comparaciones y criterios adoptados, hasta llegar a los resultados que tenemos hasta ahora.

1. Introducción

La idea de este informe es concentrar los análisis que hicimos desde comienzos de febrero hasta la fecha con los datos de turbidez del ECO FLNTU. Los datos que usamos son principalmente los de la campaña del 2019-12-17, que son los más completos y los que mejor pueden compararse de las dos campañas que realizamos con este instrumento.

Recordemos que en la campaña anterior (2019-12-10), tuvimos inconvenientes tanto con el ECO como con el OBS501 (2016), que es el instrumento que usamos como testigo para caracterizar el ECO. Hasta la fecha, estas son las únicas dos campañas que realizamos con el nuevo instrumento.

En la primera campaña, tuvimos inconvenientes para iniciar el OBS, lo que hizo que no pudieran tomarse mediciones para las primeras dos estaciones. Por otra parte, cometimos un error en la programación del ECO FLNTU antes de la campaña que causó que no tuviéramos datos para las últimas estaciones de esa campaña. Veremos que estos dos inconvenientes hicieron que esta campaña no sirviera para la caracterización del ECO.

Por este motivo, los datos utilizados de aquí en más para el análisis, corresponden principalmente a los de la campaña del 2020-12-17.

2. Discusión y resultados

2019-12-10

Teniendo en cuenta que lo que nos interesa es caracterizar el ECO FLNTU, las comparaciones que nos interesa hacer son aquellas en las que intervienen las me-

diciones llevadas a cabo con el HACH, las mediciones de *side-scattering* del OBS501-2016 y, lógicamente, el ECO FLNTU.

En primer lugar, en la figura 1 mostramos los datos correspondientes a la campaña del 2019-12-10.

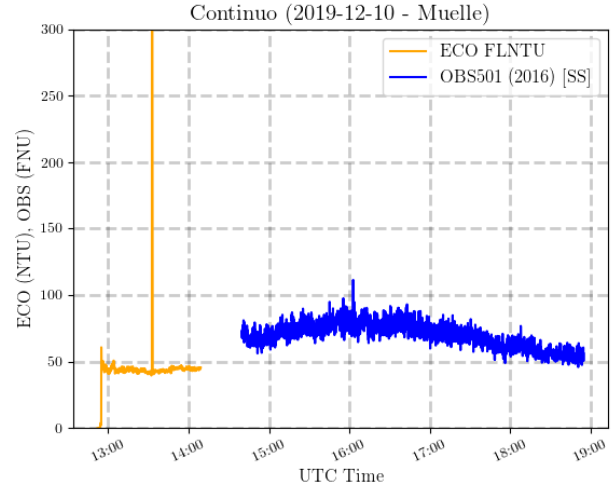


Figura 1: Datos en continuo del ECO FLNTU y el OBS501-2016 para la campaña del 2019-12-10.

Lo que se ve es que, como mencionamos antes, el OBS presentó problemas para iniciar las mediciones, mientras que el ECO las terminó de manera temprana. Lo que se ve en la figura 1 es que incluso no tenemos superposición de datos correspondientes a ambos instrumentos en un mismo momento. Es precisamente por este motivo que decidimos no considerar esta campaña para la caracterización.

Para ver el resultado que tuvo este inconveniente en términos de las estaciones de esa campaña, en la figura

2 vemos los datos resultantes para cada una.

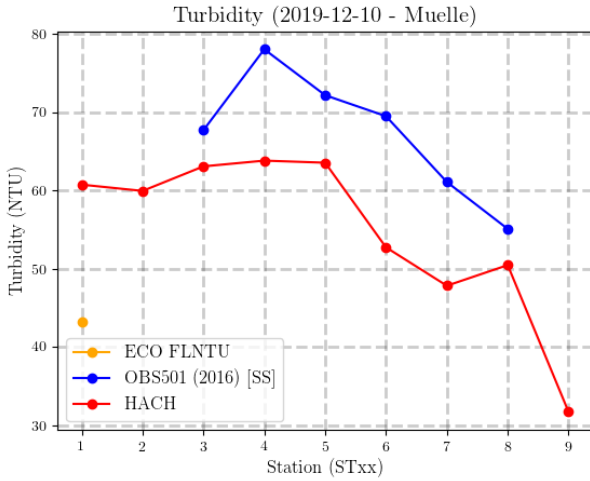


Figura 2: Datos de las nueve estaciones de la campaña del 2019-12-10.

Como vemos, nos quedó un único dato para el ECO FLNTU (la estación ST01), de la que además, únicamente tenemos datos del HACH. Esto hace que no podamos analizar ni siquiera si el instrumento sigue la tendencia del resto.

Por este motivo, decidimos centrar la caracterización en los datos de la siguiente campaña.

2019-12-17

Para la siguiente campaña (2019-12-17), en cambio, sí tenemos datos completos de todos los instrumentos.

Datos crudos

Antes de comenzar, vamos a analizar los datos crudos correspondientes al ECO y el OBS para esta campaña,. Estos datos se muestran en la figura 3.

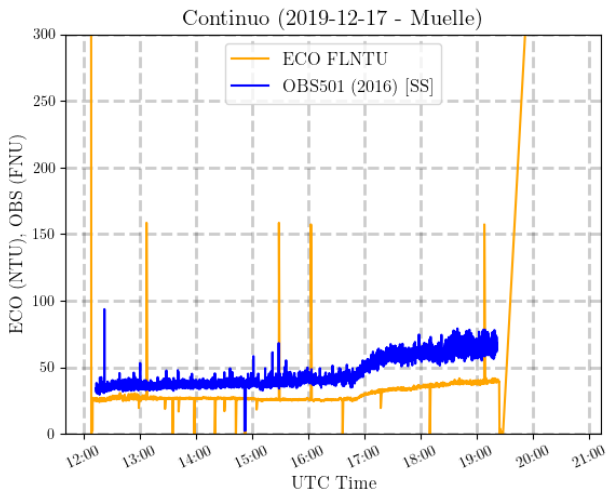


Figura 3: Datos del ECO y el OBS en continuo.

Es importante mencionar en este punto que, si bien en el caso del OBS501 los datos se muestran tal como lo entrega el instrumento, los del ECO FLNTU pasaron por el filtro de errores de transferencia.

Es importante hacer un breve análisis de lo que se ve en este gráfico. La curva graficada en amarillo corresponde a los datos del ECO FLNTU y es la que nos interesa estudiar.

El pico que se ve a las 12:08 corresponde con el encendido del instrumento, momento en que la compuerta se encontraba cerrada y por lo tanto, la turbidez medida saturaba la medición. Al abrirse la compuerta, la medición se derrumba a cero debido a que el instrumento mide en aire hasta ser sumergido. Lo mismo ocurre, en orden inverso, a las 19:27, cuando apagamos el sensor.

Tenemos también una medición de turbidez cercana a cero a las 14:53, momento en que retiramos tanto el ECO como el OBS del agua para verificar que que el ECO estuviera midiendo correctamente y no volviera a ocurrir el problema de la campaña anterior.

En cuanto al resto de los datos, lo primero que se observa es una gran cantidad de *spikes* en los datos del ECO FLNTU, que adjudicamos a errores de transferencia que pasaron el primer filtrado. Dejando esto de lado, vemos una mayor cantidad de ruido en los datos del OBS en comparación con los del ECO, esto es, sin considerar los spikes. Esto es de esperarse si tenemos en cuenta que cada medición del ECO es el promedio de 300 mediciones independientes, mientras que las del OBS vienen de una única medición.

Teniendo en cuenta que los métodos de medición de ambos instrumentos son comparables nos parece razonable asumir que los datos provenientes de ambos instrumentos deberían tener una amplitud de ruido del mismo orden. Esto significa que la relación ruido/señal del ECO debería ser $\sqrt{300} \sim 17$ veces más pequeño que el del OBS.

Ahora nos queda solamente mostrar que los spikes no provienen de mediciones sino de errores que se producen en la transferencia de los datos del instrumento hacia la computadora.

Afortunadamente, de esta segunda campaña de diciembre tenemos dos transferencias independientes, a las que llamamos «transferencia A» y «transferencia B». En la figura 4 vemos estos datos.

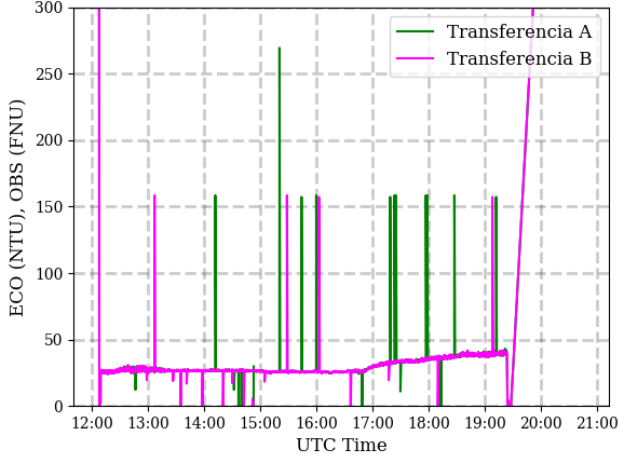


Figura 4: Comparación de los datos de los mismos datos, transferidos dos veces a la computadora.

Lo que vemos es que los spikes no se repiten en ambos conjuntos de datos, lo que significa que son errores que se producen en el momento de la transferencia de datos y no en el momento de la medición. Esto es un buen resultado porque indica que los datos están sanos y que podrían recuperarse en caso de ser necesario, realizando múltiples transferencias.

Remoción de outliers

Para remover los outliers, utilizamos el mismo método que se utiliza normalmente para los datos del OBS501.

El método consiste en obtener los cuartiles Q_1 y Q_3 y luego calcular el rango intercuartil IQR como:

$$IQR = Q_3 - Q_1$$

El siguiente paso entonces es eliminar cualquier dato que se encuentre por debajo de $Q_1 - 1,5 \times IQR$ o por encima de $Q_3 + 1,5 \times IQR$.

Los resultados de aplicar este método sobre los datos crudos se muestra en la figura 5.

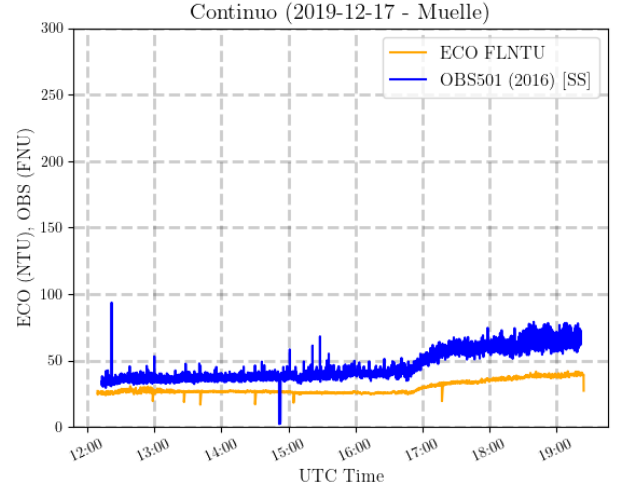


Figura 5: Datos del ECO y el OBS sin outliers.

Lo que vemos es que este método funciona satisfactoriamente para la remoción de los outliers presentes en los datos del ECO, ya sea los spikes causados por errores de transferencia como los momentos en que se midió fuera del agua, en ambos extremos del gráfico.

Más aún, retomando la cuestión de las amplitudes relativas del ruido en las señales de los dos instrumentos, debido a que el ECO FLNTU almacena los promedios de 300 mediciones, vemos que este método remueve los outliers de manera mucho más eficiente que en el OBS.

Para ver mejor esto, en la figura 6 vemos el detalle de las 14:53, en que retiramos ambos instrumentos del agua.

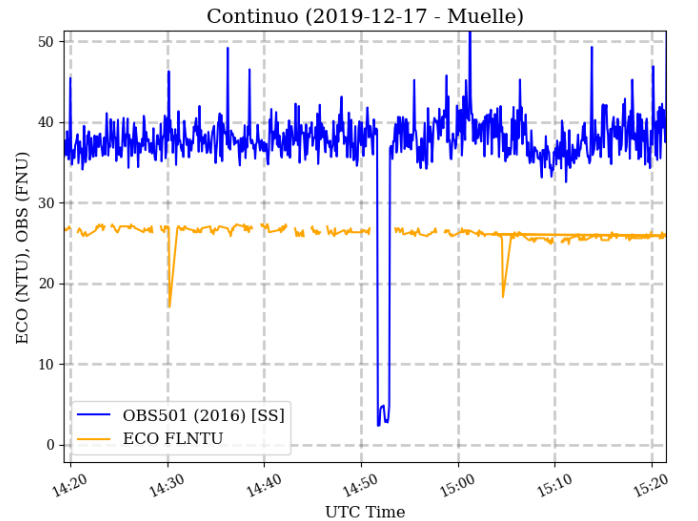


Figura 6: Detalle de las 14:53, en que retiramos los instrumentos del agua.

Lo que vemos es que el filtrado de outliers alcanzó incluso para remover los datos correspondientes a los instantes en que los instrumentos estuvieron registrando mediciones sin estar sumergidos, lo que no ocurrió para

los datos del OBS.

Suavizado de las curvas

Finalmente, sobre los datos ya limpiados de errores de transferencia y sin outliers, realizamos los suavizados de las curvas, utilizando una vez más, modificaciones del método ya existente para los datos del OBS.

El método de suavizado consiste en tomar, para cada medición, el promedio de todas las mediciones realizadas durante la ventana de un minuto centrada en ella. En otras palabras, lo que se hace es promediar la medición junto con las mediciones que la preceden hasta 30 segundos, y las que la suceden hasta 30 segundos.

En la figura 7 vemos el gráfico de las mediciones en continuo de ambos instrumentos junto con las suavizadas, manteniendo la misma escala que en los anteriores.

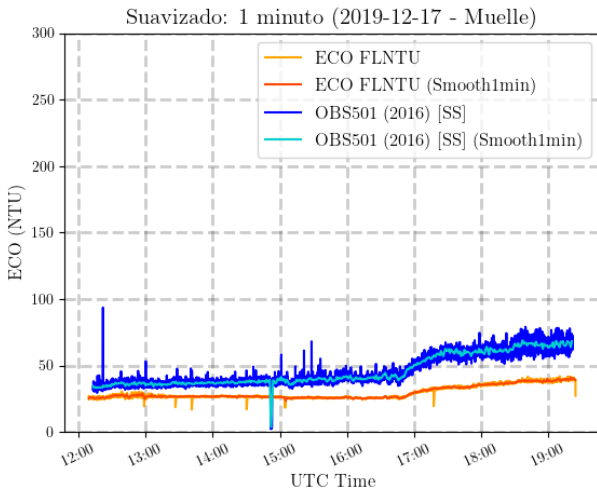


Figura 7: Datos del ECO y el OBS: mediciones en continuo y suavizadas.

Lo que se ve es lo esperable de una suavización de los datos: la disminución en el ruido en ambas señales.

Para ver mejor el detalle de los datos, en la figura 8 relajamos la escala nuevamente y mostramos únicamente los datos suavizados, manteniendo el código de colores del gráfico anterior.

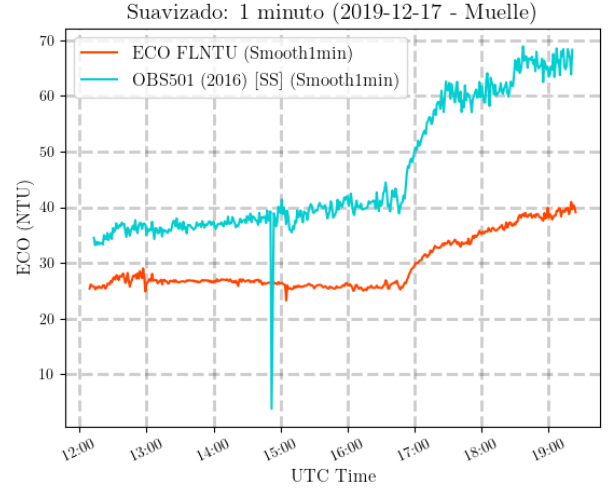


Figura 8: ECO y OBS: únicamente datos suavizados.

Valores de las estaciones

Con los datos ya suavizados, es interesante ver si los tres instrumentos utilizados en la campaña para medir turbidez (HACH, OBS501 y ECO FLNTU) presentan comportamientos similares.

Dado que el HACH no mide de manera continua, sino que se utiliza manualmente para tomar mediciones en cada estación, es razonable realizar comparaciones con los valores correspondientes a cada una de las doce estaciones de esta campaña. Se muestra esta comparación en la figura 9.

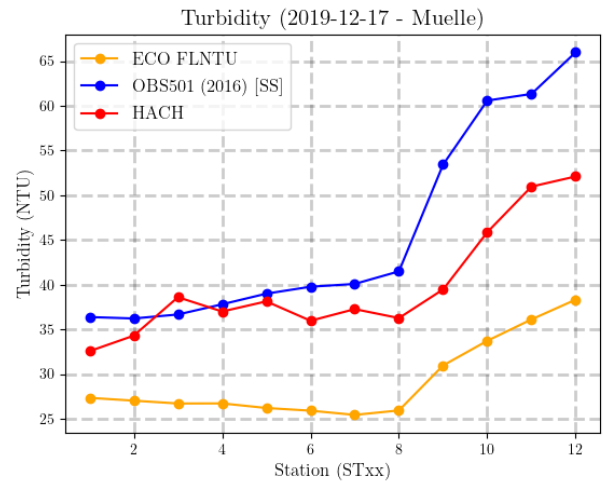


Figura 9: Valores finales del ECO FLNTU, el OBS501 y el HACH para las doce estaciones.

El eje x corresponde a las doce estaciones de la campaña, cuyas etiquetas son: {ST01, ST02, ..., ST12}.

Lo que se ve es que los tres instrumentos presentan las mismas tendencias a lo largo de la campaña, lo que da el indicio de que la hipótesis de proporcionalidad entre los valores medidos por cada uno de ellos es correcta.

La idea es entonces comparar el ECO FLNTU con los otros dos instrumentos, para encontrar las constantes de proporcionalidad.

Comparamos primero el ECO con el OBS501 y realizamos un ajuste lineal sobre los datos. Los resultados pueden verse en la figura 10.

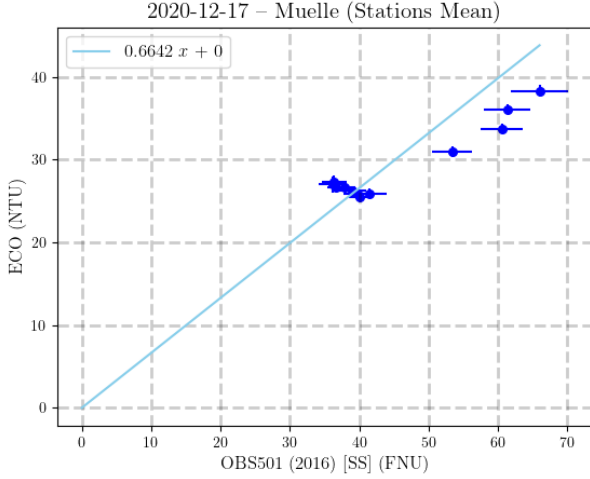


Figura 10: Ajuste lineal sobre las mediciones de turbidez del ECO FLNTU para cada campaña en función de las del OBS501-2016.

El ajuste arroja un valor de pendiente de 0,6642 y un valor de chi cuadrado reducido de $\chi^2_\nu \sim 3,80$.

Por otra parte, realizamos el ajuste análogo sobre los datos del ECO FLNTU en función de los del HACH. Estos resultados pueden verse en la figura 11.

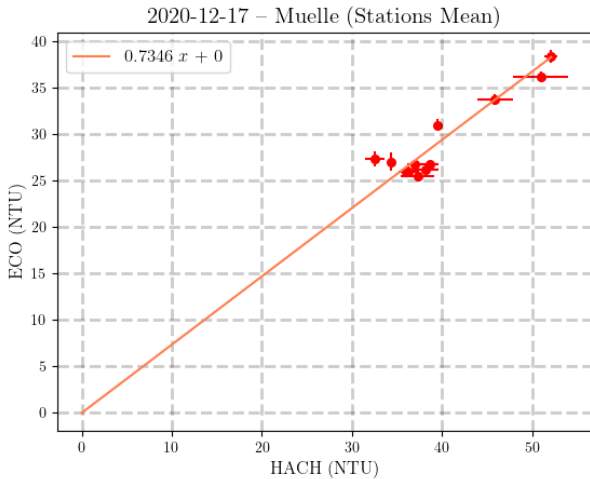


Figura 11: Ajuste lineal sobre las mediciones de turbidez del ECO FLNTU para cada campaña en función de las del HACH.

Para este caso, El ajuste arroja un valor de pendiente de 0,7345 y un valor de chi cuadrado reducido de $\chi^2_\nu \sim 2,95$.

Para ambos ajustes se fijó la coordenada al origen en cero y se utilizó el método ODR (Orthogonal Distance Regression), de modo de contemplar las incertezas en las dos magnitudes.

Ajuste sobre las mediciones suavizadas del ECO FLNTU y OBS501-2016

Para terminar, y para dar una idea más completa de lo discutido recién, veamos el ajuste realizado sobre los datos del ECO FLNTU en función de los datos del OBS501-2016, pero realizados sobre todos los datos en continuo, y no solamente los correspondientes a las estaciones de la campaña.

Para un mejor resultado, utilizamos los datos suavizados en ventanas de un minuto, de modo de reducir la cantidad de outliers en los datos (en especial, los del OBS, que los presenta en mayor cantidad). El resultado puede verse en la figura 12.

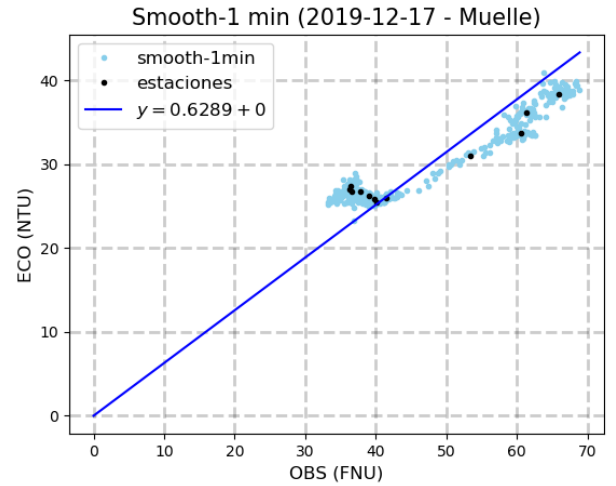


Figura 12: Ajuste sobre los datos suavizados (1 minuto). En azul, los puntos ajustados, correspondientes a los datos suavizados, y en negro, los valores de las estaciones.

Los puntos celestes corresponden a los datos suavizados y son los que se utilizaron para llevar a cabo el ajuste lineal y los puntos negros son los valores obtenidos para las estaciones, que no intervinieron en el ajuste.

Nuevamente, se fijó la coordenada al origen en cero y se utilizó el método ODR para contemplar los errores de ambos instrumentos. El resultado para la pendiente (0.6289) es consistente con el obtenido en el ajuste solo utilizando los valores correspondientes a las estaciones (0.6642), siendo que la diferencia las pendientes es aproximadamente de un 5 %.

3. Conclusiones

Luego de haber visto los resultados de la campaña del 2019-12-17, consideramos que los datos obtenidos con el ECO FLNTU son satisfactorios ya que, como vimos, los datos se ajustan a lo que esperábamos en cada uno de los análisis realizados.

Vimos que fue posible resolver los problemas que aparecieron en la transferencia de los datos mediante los procesos de filtrado y suavizado.

Tal como esperábamos, encontramos que el ECO FLNTU produce datos consistentes con los instrumentos ya conocidos, siendo que los datos arrojados por todos ellos efectivamente resultan proporcionales entre sí.

Por otra parte, observamos que los datos del ECO FLNTU presentan una relación ruido/señal considerablemente menor que el OBS501. Esto puede verse no solo en la cantidad de spikes presente en cada conjunto de datos, sino también en la facilidad para eliminar los eventos en que los instrumentos fueron retirados del agua utilizando solamente la el método de remoción de outliers.

Teniendo en cuenta estos aspectos, consideramos que tanto los controles de la transferencia como los de detección de anomalías temporales y globales dan como resultado datos apropiados para estudios futuros.