**Documento explicativo para el procesado de los Langleys. Métodos Brewer / Dobson.**

Se explica en este documento cual es el proceso y cuales las funciones involucradas en el análisis de los Langleys plots para el Brewer. Las funciones involucradas en el proceso se hallan todas bajo el directorio matlab/Langley/Langley\_dev.

Se trabaja con dos metodologías ligeramente diferentes, correspondientes a los usados en las redes Brewer y Dobson.

* Brewer Algorithm :

Siendo , donde (βi Rayleigh coefficient, m3 ozone airmass y m2 Rayleigh airmass). Si asumimos Ozono constante, entonces podremos hacer ,  es decir, la ETC será el corte con el eje de ordenadas para m3=0.

* Dobson algorithm: En este caso se aplica la estrategia ***1/m3*** usada por los Dobsons, resultando

Lo que se obtiene ahora, a través de la pendiente ETCc, es la corrección a la ETC **estimada** ETCo. Las ventajas del método Dobson frente al método Brewer serían:

1. Las observaciones están mejor distribuidas frente a 1/m3 que frente a m3.
2. El error cometido al determinar la pendiente es inferior al error cometido al determinar el corte en m3=0.

En principio trabajaremos con las medidas sincronizadas entre los instrumentos. El tiempo de sincronización es Tsync = 5 minutos, siendo este parámetro definido *hardcode* en todos los scripts, es decir, en caso de querer modificarlo, habría que hacerlo directamente en el código. Puesto que se ha respetado igual estructura para las matrices de datos independientemente de que datos son los usados se podrá trabajar bien con los ***sumarios*** bien con las ***medidas individuales*** (5 por cada sumario). Notar que en el caso de las medidas individuales se puede trabajar igualmente con medidas simultáneas (ahora será Tsync=1/2 minuto).

Tabla : Sumario de variables devueltas por la función read\_bdata.m

|  |  |
| --- | --- |
| **ozone.dsum** | Salidas raw, tomadas directamente del fichero B (sumarios) -> B config (1). Ratios. |
| **ozone.raw0** | Salidas raw, tomadas directamente del fichero B (med. individuales) -> B config (1). Ratios & counts recalc. |
| **ozone.ozone\_sum** | Summaries, recalculated (1 & 2 configs). No MS9`s, No raw counts |
| **ozone.ozone\_ds** | Ind. measurements, recalculated (1 & 2 configs). Ratios & MS9’s |

Tenemos entonces dos posibilidades:

* Sumarios Simultáneos.

La función en este caso sería **Langley\_summ\_sync**, recibiendo como *inputs* los siguientes: summary, summary\_old y variable de definiciones Cal.Vemos que para trabajar de esta manera se asume que contamos con los sumarios ya generados, es decir, habrá que seguir el procedimiento ordinario ***read\_bdata -> sl\_report\_jday -> read\_cal\_config\_new -> test\_recalculation -> filter\_corr (opcional)***.

No voy a entrar en detalles acerca de cómo obtener las variables summary y summary\_old, pero si conviene tener en cuenta lo siguiente:

* **Aplicamos la función filter\_corr.m a los sumarios**. Entonces tendremos MS9 con y sin corrección por filtros, campos 8 y 9 de summary, respectivamente.
* **No aplicamos la función filter\_corr.m a los sumarios**. Entonces tendremos MS9 sin corrección por filtros (campos 8). El campo 9 es ahora la desviación estándar

Por defecto se trabaja con el campo 8, de modo que la forma de trabajo preferida será por defecto aplicar siempre la función filter\_corr.m, y entonces trabajar bien con ***summary***/***summary\_old*** (filter corrected) bien con ***summary\_orig***/***summary\_orig\_old*** (no filter corrected).

La salida de la función **langley\_summ\_sync** será un cell array para cada instrumento (notar que sólo cabe trabajar con múltiples instrumentos, pues buscamos medidas simultáneas). Cada una de estas celdas contiene a su vez una celda para cada día conteniendo una matriz de datos de 39 columnas con la siguiente estructura:

Tabla 2: Salida de la función langley\_summ\_sync.m

|  |
| --- |
| ***time, lat, long, sza, m3, m2, flag,  NaN, tst, filt, temp*** ... % 1-11  ***NaN, NaN, NaN, NaN, NaN, NaN, NaN***  ... % 12-18 c/s recalculated 1 (Rayleigh uncorr.!!)  ***O3 old, NaN, NaN, NaN, NaN, NaN, MS9 old***  … % 19-25 ratios recalculated 1 (Rayleigh corr.!!)  ***NaN, NaN, NaN, NaN, NaN, NaN, NaN*** ... % 26-32 c/s recalculated 2 (Rayleigh uncorr.!!)  ***O3 old, NaN, NaN, NaN, NaN, NaN, MS9 old***  ... % 33-39 ratios recalculated 2 (Rayleigh corr.!!) |

La ventaja de trabajar con los sumarios reside, entre otras, en que además podemos usar el ozono para descartar días en los que este parámetro no sea constante.

* Medidas Individuales Simultáneas.

No es directo obtener una variable con los sumarios que contenga las ratios MS9 y además las cuentas recalculadas (o brutas) para cada canal. La alternativa será calcular a partir de las medidas individuales los sumarios (lo que se usa para dos cosas: 1) depurar medidas de ozono, 2) obtener el ozono), y a partir de aquí obtener las medidas simultáneas (ahora si dispondríamos de toda la información que se quiera: cuentas brutas, cuentas recalculadas, ratios, ozono, …). Trabajamos en este caso con la función **langley\_data\_cell.m**, recibiendo como *inputs* las variables ***ozone\_raw, ozone\_ds*** y ***config*** (salidas de readb\_data.m), y obteniendo como *output*:

* ***ozone\_lgl***: serían las medidas individuales (5 por cada sumario), depuradas según el procedimiento estándar usado con los sumarios (normal O3\_std < 2.5, 100<O3<600, hg flag =1, n = 5).
* ***ozone\_lgl\_sum***: sumarios recalculados a partir de las medidas individuales anteriores. Los campos disponibles son los detallados en la tabla siguiente. Obtendremos esta variable sólo en caso de pedirla.

Tabla 3: Salida de la función langley\_data\_cell.m: variables ozone\_lgl y ozone\_lgl\_sum

|  |
| --- |
| ***'date' 'hg\_id' 'nds' 'sza' 'm2' 'm3' 'sza' 'saz' 'tst' 'filt' 'temp'*** *...% 1-11*  ***'f0' 'f1' 'f2' 'f3' 'f4' 'f5' 'f6' ...*** *% 12-18 count-rates recalculated 1 (Rayleigh uncorrected !!)*  ***'o3\_1' 'r1' 'r2' 'r3' 'r4' 'r5' 'r6' ...*** *% 19-25 ratios recalculated 1 (Rayleigh corrected !!)*  ***'F0' 'F1' 'F2' 'F3' 'F4' 'F5' 'F6' ...*** *% 26-32 count-rates recalculated 2 (Rayleigh uncorrected !!)*  ***'O3\_2' 'R1' 'R2' 'R3' 'R4' 'R5' 'R6' ...*** *% 33-39 ratios recalculated 2 (Rayleigh corrected !!)* |

* ***cfg***: Constantes de calibración en juego. Estructura con campos old y new.

Tabla : Salida de la función langley\_data\_cell.m: variable cfg

|  |
| --- |
| ***'Usage date', 'o3 Temp coef 1', 'o3 Temp coef 2', 'o3 Temp coef 3', 'o3 Temp coef 4', 'o3 Temp coef 5', 'O3 on O3 Ratio', 'ETC on O3 Ratio', 'Dead time (sec)', 'ND filter 0', 'ND filter 1', 'ND filter 2', 'ND filter 3', 'ND filter 4', 'ND filter 5'*** |

* ***lgl\_leg***: legenda para cada una de las variables de salida

Como se ve, la estructura de los datos simultáneos (sumarios) y de los individuales es idéntica. Esto nos permitirá desarrollar las funciones de análisis de forma que sean comunes para cualquier tipo de dato de entrada.

En la figura abajo se muestran como ejemplo los resultados de los Langleys obtenidos de los sumarios calculados a partir de la función test\_recalculation.m (puntos) y los obtenidos de los sumarios recalculados a partir de las medidas individuales, función langley\_data\_cell.m (círculos). Como se muestra, los resultados en ambos casos son idénticos, salvo pequeñas diferencias debidas a:

* Los resultados obtenidos a partir de la variable summary no están filtrados de ninguna manera.
* En el caso de los sumarios obtenidos a partir de las medidas individuales se aplica por defecto el siguiente filtro: las medidas no será válidas al menos que como mínimo se disponga de 100 ds’s (es decir, 20 sumarios).
* En el caso de los sumarios obtenidos a partir de las medidas individuales no estamos trabajando con medidas simultáneas.

General__ETC_all_157.tif

Figure 1: Resultado del análisis de los Langley Plots sobre dos conjuntos de datos diferentes: a) MS9 simultáneas para la triada, obtenidas a partir de los sumarios recalculados por la función test\_recalculation.m (black, AM, and blue, PM, dots) y b) MS9 obtenidas a partir de los sumarios recalculados directamente a partir de las variables ozone\_raw y ozone\_ds (black, AM, and blue, PM, circles). En este último caso no se trata de medidas simultáneas. Métodos Brewer (arriba) y Dobson (abajo).

Paso a comentar a continuación primero la depuración de los datos y luego el análisis de los Langleys:

1. **Depuración de los datos, nivel 1 (langley\_filter\_lvl1.m)**

La función langley\_filter\_lvl1.m recibe como input los datos de langley, bien los datos simultáneos obtenidos a partir de los sumarios bien las medidas individuales. Funciona con ambas. Los input opcionales se refieren a la depuración que queramos llevar a cabo, que por otra parte se aplica a “medios-días”, AM-PM, definidos a partir del tiempo solar verdadero. Son los siguientes (notar que estos incluyen a los que ya se definió en su día como Data Selection Level1):

* ***airmass:*** rango de masas ópticas de trabajo. Puede ser un solo valor ó una pareja de valores. No filtro ([]) por defecto.
* ***N\_filt:*** número mínimo de medidas por filtro. 5 por defecto (NOT IMPLEMENTED).
* ***N\_hday:*** número mínimo de medidas para cada “medio-día”. 20 sumarios → 20 x 5 medidas individuales, por defecto.
* ***O3\_hday:*** máxima desviación estándar para el ozono promedio de cada “medio-día”. Se pretende con este filtro descartar los “medios-días” en los que el ozono no permanece aproximadamente constante. Por defecto No se usa este filtro.
* ***Date\_range:*** Rango de fechas a analizar. No filtro ([]) por defecto.
* ***F\_corr:*** Corrección a los filtros de atenuación. Pueden ser de dos formas: variable F\_corr generada por read\_cal\_config\_new. Esta sería la forma preferida (1 vector / day). 6-D vector (e.g. [0 0 0 11 NaN 0]).
* ***AOD:*** AOD lvl1.5 obtenido de la página web de AERONET. No filtro por defecto.
* ***Summ:*** Argumento opcional necesario para poder trabajar con sumarios. Los valores posibles son 0 (no sumarios) ó 1 (sumarios). 0 por defecto. Este es un input adicional que, aunque opcional, es importante definir explícitamente, o al menos tener presente su significado. Como hemos dicho, la función langley\_filter\_lvl1.m admite como entrada bien los datos de langley obtenidos a partir de los sumarios (medidas simultáneas), o bien los datos de langley obtenidos a partir de las medidas individuales. En este último caso, se ofrecen dos opciones posibles: 1) medidas individuales y 2) sumarios calculados a partir de esas medidas individuales. Existe entonces un filtro adicional (digamos level 0) que no es más que el fitro rutinario aplicado a los sumarios, es decir O3\_std < 2.5, 100<O3<600, hg flag =1, n = 5. En el caso de que estemos trabajando con la variable ***langsumm\_sync\_data*** entonces este filtro inicial no es necesario, ya ha sido implementado, pero en el caso de que trabajemos bien con la variable ***ozone\_lgl*** bien con ***ozone\_lgl\_sum***, entonces debemos aplicarlos. Si se trata de la primera (***ozone\_lgl***), asignamos a ‘summ’ el valor 0, y lo que obtenemos son las medidas individuales pero se han descartado aquellas cuyo sumario asociado no cumpla con las condiciones anteriores. Si se trata de la segunda (***ozone\_lgl\_sum***) entonces asignamos a ‘summ’ el valor 1, y se aplicará el filtro de los sumarios. En el caso de la variable ***langsumm\_sync\_data*** también debemos hacer ‘summ ‘ = 1, sólo que ahora no pasará nada. En definitiva, siempre que trabajamos con sumarios debemos iniciar el valor del argumento opcional ‘summ’ a 1.
* ***Plots:*** En caso de definirlo a 1, nos dará un ploteo para cada medio-día mostrando la relación MS9-airmass junto con el ozono, y los residuos del ajuste. Se muestra un ejemplo en la figura abajo.

General__chk_185.tif

Figure :

Como salida, la función langley\_filter\_lvl1.m devuelve una matriz de datos con igual estructura que la de entrada, pero depurada según los criterios anteriores. Además, nos devuelve en pantalla una tabla con los diferentes parámetros asociados a cada “medios-días”, como se muestra en la tabla siguiente:

Tabla 6: Parámetros asociados a cada día analizado.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Diaj** | **AM** | **PM** | **O3\_std (am)** | **N(am)** | **O3\_std (pm)** | **N(pm)** | **AOD\_340** |
| **01-Sep-2013** | 244 | 1 | 1 | 1.57 | 58 | 1.72 | 60 | 0.093 |
| **02-Sep-2013** | 245 | 1 | 1 | 1.45 | 60 | 1.83 | 57 | 0.032 |
| **03-Sep-2013** | 246 | 1 | 1 | 1.32 | 36 | 1.49 | 27 | 0.028 |
| **04-Sep-2013** | 247 | 1 | 1 | 1.71 | 60 | 1.84 | 62 | 0.033 |
| **05-Sep-2013** | 248 | 1 | 1 | 2.12 | 60 | 1.60 | 53 | 0.094 |
| **06-Sep-2013** | 249 | 1 | 1 | 2.00 | 49 | 3.34 | 43 | 0.154 |
| **07-Sep-2013** | 250 | 1 | 1 | 2.54 | 50 | 2.01 | 33 | 0.015 |
| **08-Sep-2013** | 251 | 1 | 1 | 1.85 | 46 | 1.82 | 41 | 0.016 |
| **09-Sep-2013** | 252 | 1 | 1 | 2.06 | 53 | 1.69 | 66 | 0.016 |
| **10-Sep-2013** | 253 | 1 | 1 | 1.65 | 53 | 2.24 | 64 | 0.038 |
| **11-Sep-2013** | 254 | 1 | 1 | 1.99 | 55 | 1.75 | 55 | 0.072 |
| **12-Sep-2013** | 255 | 0 | 0 | 0.87 | 4 | 1.55 | 1 | 0.021 |
| **13-Sep-2013** | 256 | 1 | 1 | 1.90 | 44 | 2.24 | 62 | 0.010 |
| **14-Sep-2013** | 257 | 1 | 1 | 1.71 | 53 | 1.85 | 61 | 0.014 |
| **15-Sep-2013** | 258 | 1 | 1 | 1.84 | 49 | 1.79 | 59 | 0.016 |
| **16-Sep-2013** | 259 | 1 | 1 | 1.87 | 44 | 1.85 | 51 | 0.031 |
| **17-Sep-2013** | 260 | 1 | 1 | 1.93 | 42 | 2.03 | 51 | 0.012 |
| **18-Sep-2013** | 261 | 1 | 1 | 1.58 | 58 | 1.74 | 62 | 0.010 |
| **19-Sep-2013** | 262 | 1 | 1 | 1.77 | 20 | 1.72 | 62 | 0.004 |
| **20-Sep-2013** | 263 | 1 | 1 | 1.65 | 52 | 1.85 | 59 | 0.007 |
| **21-Sep-2013** | 264 | 1 | 1 | 1.96 | 53 | 1.99 | 58 | 0.012 |
| **22-Sep-2013** | 265 | 1 | 1 | 1.91 | 53 | 4.15 | 62 | 0.012 |
| **23-Sep-2013** | 266 | 1 | 1 | 1.50 | 57 | 1.74 | 60 | 0.007 |
| **24-Sep-2013** | 267 | 1 | 1 | 1.70 | 53 | 2.11 | 61 | 0.007 |
| **25-Sep-2013** | 268 | 1 | 1 | 2.30 | 54 | 1.68 | 56 | 0.009 |
| **26-Sep-2013** | 269 | 1 | 1 | 2.12 | 55 | 1.67 | 55 | 0.014 |
| **27-Sep-2013** | 270 | 1 | 1 | 2.35 | 48 | 1.34 | 21 | 0.014 |
| **28-Sep-2013** | 271 | 1 | 1 | 1.79 | 56 | 1.87 | 46 | 0.012 |
| **29-Sep-2013** | 272 | 1 | 1 | 1.56 | 57 | 1.60 | 59 | 0.011 |
| **30-Sep-2013** | 273 | 1 | 1 | 1.59 | 54 | 1.77 | 61 | 0.006 |
| **01-Oct-2013** | 274 | 1 | 1 | 2.08 | 52 | 1.87 | 53 | 0.028 |
| **02-Oct-2013** | 275 | 0 | 0 | 1.43 | 17 | 1.79 | 6 | 0.018 |
| **04-Oct-2013** | 277 | 1 | 0 | 1.33 | 23 | 1.56 | 20 | 0.014 |
| **05-Oct-2013** | 278 | 1 | 1 | 2.34 | 50 | 1.81 | 56 | 0.013 |
| **06-Oct-2013** | 279 | 1 | 1 | 1.68 | 51 | 1.60 | 55 | 0.015 |
| **07-Oct-2013** | 280 | 1 | 1 | 1.77 | 51 | 1.61 | 58 | 0.009 |
| **08-Oct-2013** | 281 | 1 | 1 | 1.52 | 51 | 1.42 | 59 | 0.009 |
| **09-Oct-2013** | 282 | 1 | 1 | 1.54 | 54 | 1.69 | 60 | 0.008 |
| **10-Oct-2013** | 283 | 1 | 1 | 2.64 | 55 | 1.97 | 55 | 0.010 |
| **11-Oct-2013** | 284 | 1 | 1 | 1.78 | 53 | 1.75 | 54 | 0.011 |
| **12-Oct-2013** | 285 | 1 | 1 | 1.89 | 52 | 1.51 | 55 | 0.013 |
| **13-Oct-2013** | 286 | 1 | 1 | 1.57 | 53 | 1.62 | 54 | 0.015 |
| **14-Oct-2013** | 287 | 1 | 1 | 1.84 | 53 | 1.56 | 57 | 0.010 |
| **15-Oct-2013** | 288 | 1 | 1 | 1.44 | 49 | 1.53 | 59 | 0.014 |
| **16-Oct-2013** | 289 | 1 | 1 | 1.43 | 56 | 1.54 | 55 | 0.016 |
| **17-Oct-2013** | 290 | 1 | 1 | 2.77 | 54 | 1.62 | 59 | 0.040 |

1. **Obtención de la constante ETC mediante Langley plot (langley\_analys.m)**

La función langley\_analys.m realiza una regresión simple de MS9 frente a la masa óptica para obtener la constante ETC. Aunque no está completa, los cálculos básicos si están implementados.

Los argumentos de entrada son, como obligatorios, los siguientes:

1. **lgl\_data:** puede ser o bien la salida depurada de la función langley\_filter\_lvl1.m o bien cualquiera de las variables (no depuradas) ***langsumm\_sync\_data,*** ***ozone\_lgl*** ó ***ozone\_lgl\_sum.***
2. **brw:** puesto que los datos de entrada van a ser cell-arrays, uno para cada instrumento, con este parámetro identificamos el instrumento que nos interesa estudiar.
3. **Cal:** esta es una variable-estructura normalmente resultado de la evaluación del fichero de definiciones (por ejemplo calizo2013\_setup.m). En todo caso de aquí sólo se toman las configuraciones 1 y 2.

|  |
| --- |
| General_.tif |
| General_.tif |

Figure 3: Efecto del filtrado adicional ‘res\_filt’. Se eliminan los datos cuyo residuo sea mayor que 1.5 veces la desviación estándar del total de residuos. Datos brutos (no se ha aplicado ningún otro filtro). Método Brewer (arriba) y método Dobson (abajo).

Además de los argumentos obligatorios anteriores, existe uno opcional ***‘res\_filt’.*** Este parámetro controla si se realiza o no un filtro adicional de depuración sobre los datos de la regresión, descartando aquellos que verifican la condición abs(residuos)>1.5 x std(residuos). Es decir, se eliminan aquellos datos cuyo residuo sea mayor que 1.5 veces la desviación estándar del total ([Michalsky](http://www.esrl.noaa.gov/gmd/grad/neubrew/docs/publications/Harrison_ApplOpt1994.pdf)). Una vez hecho esto, se vuelve a calcular la regresión con los nuevos datos depurados.

Como posibles salidas, hasta ahora la función devuelve las variables ***resp\_brw*** y ***resp\_dbs.*** En ambos casos ***s***e trata de matrices 3-D (:,3,2), donde:

* 1D: una fila para cada día analizado (con al menos AM o PM válidos)
* 2D: 3 columnas. Date, Langley AM y Langley PM
* 3D: 1 config / 2 config

Optional

  a) Data selected michalsky

<http://www.esrl.noaa.gov/gmd/grad/neubrew/docs/publications/Harrison_ApplOpt1994.pdf>

    same output but adding  a flag

  b) AOD from aeronet from the same period

  c) Color index

  d) Cloud observations / Cloud camera

 2) Data selection 1

    “Variables”  for selection  level 1

           b) n air mass range  -> default 1.15-3.5

           c) n observations  ( equally distributed better) -> default 25

           d) number of measurements by filter                 -> default 5

           e) Filter correction                                              -> default from config

3)    Regression  (MS9-ETCR)/n  vs 1/n ->  Linregress -> Store all the outputs of linregress  !!. In addition:  airmass range (min, max) , n obs

 output 1;  daily (AM/PM) langleyplot  (1/nu)  am/pm of the langley colors by filter  by flag

                 R6 vs time

Summary output 1:

**Simultaneous ETC correction for all the brewer per day per config**

**ETC  correction instrument vs instrument**

**ETC  correction instrument vs AOD**

4) Recalculation

        ozone mean and std recalculated with ETC corrected

5)  Filter #2  for regression results

                  n obs

                  ozone mean and std during langley

                  R coeff corr

                  Error in slope