# Instructions for generating LTA[1] Probability Bands using stylesheets

These instructions will explain how to generate the data that make up the coloured LTA bands e.g. in *Figure 1*. These bands might also be used to show maps of current flows/level/etc against their LTA ranks, e.g. in *Figure 2*.

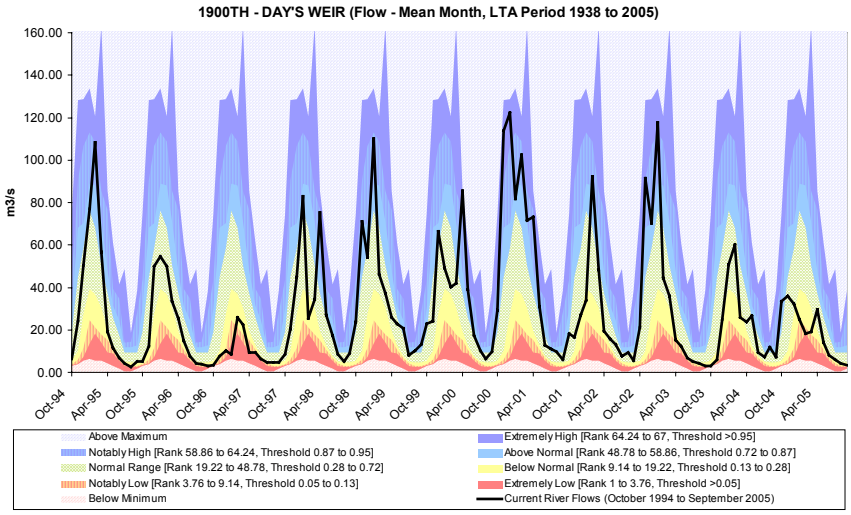



***Figure 1 – Example of a graph showing recent river flows (black line) compared to the LTA probability bands***

Generating these LTA bands is a simple process using XSL[2] stylesheets. These bands help to compare how extreme – or how normal – the situation is at a river flow, groundwater or rainfall site (etc).

There is a requirement to use these bands as part of the Agency's AMS[3] "Categorising Hydrological Data by Probability Ranking", which has been produced by the National Hydrology Policy & Process Team.

[1] LTA = Long Term Average

[2] XSL = eXtensible Stylesheet Language, which is a derivative of XML (eXtensible Mark-up Language). Stylesheets are generally used to either:

- format XML data, e.g. into a webpage or a text file format
- manipulate XML data, e.g. to perform calculations like this LTA ranking process.

[3] AMS = Environment Agency Management System

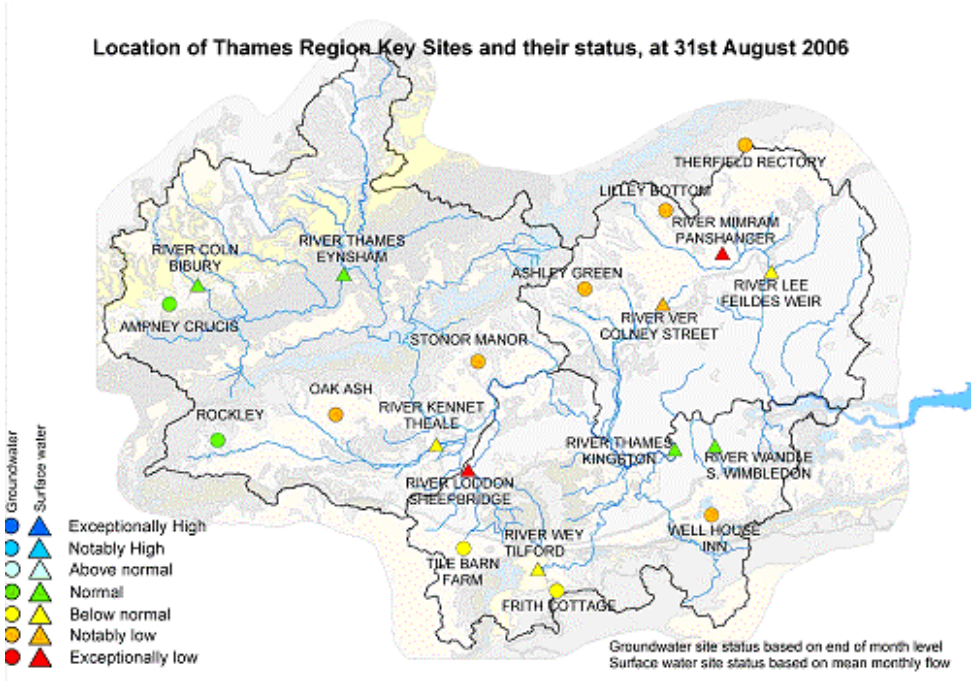



*Figure 2 – Example from the "Thames Region Monthly Hydrological Report" of a number of river/groundwater sites, which are coloured to indicate how their current flows/levels compare to the LTA probability ranks*

## Contents

## Procedure 1 – Creating LTA bands for one site

There are 5 simple steps to calculate the LTA bands:

1. Generate an XML time-series file containing the data that you want to rank[4]; for example the monthly-mean flows for the period of record at a site. You will probably be generating your data from the EA's Wiski archive, in which case the instructions in Appendix 3 may be of use. Note that you ought to make sure that the dataset is complete (see Appendix 2) and for whole years (e.g. July 1958 to June 2005).

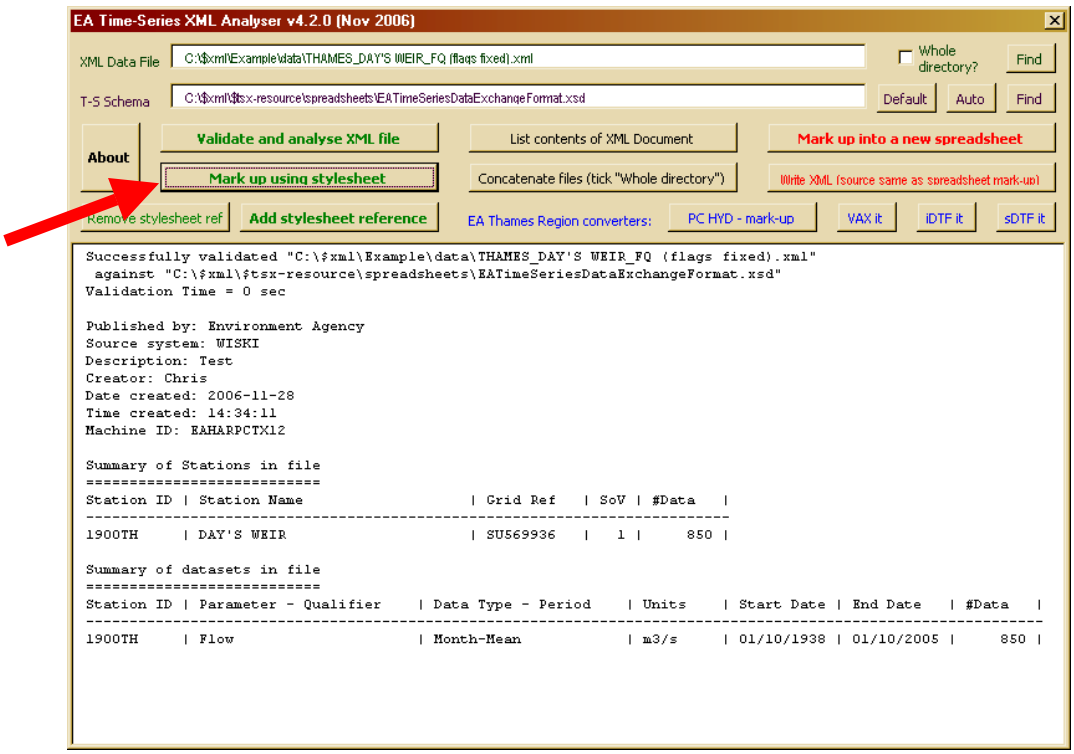



***Figure 3 – The XML Analyser***

2. Open the XML Analyser[5] spreadsheet. If you do not have a copy, you can find it at www.chrisbeales.net/tsdxs (alongside the stylesheets that you need to do this ranking process). When you open the Analyser you should see a window pop up that looks like that in *Figure 3*.

3. Click on the "Find" button, in the top-right-hand-corner of the Analyser, and locate your XML data file.

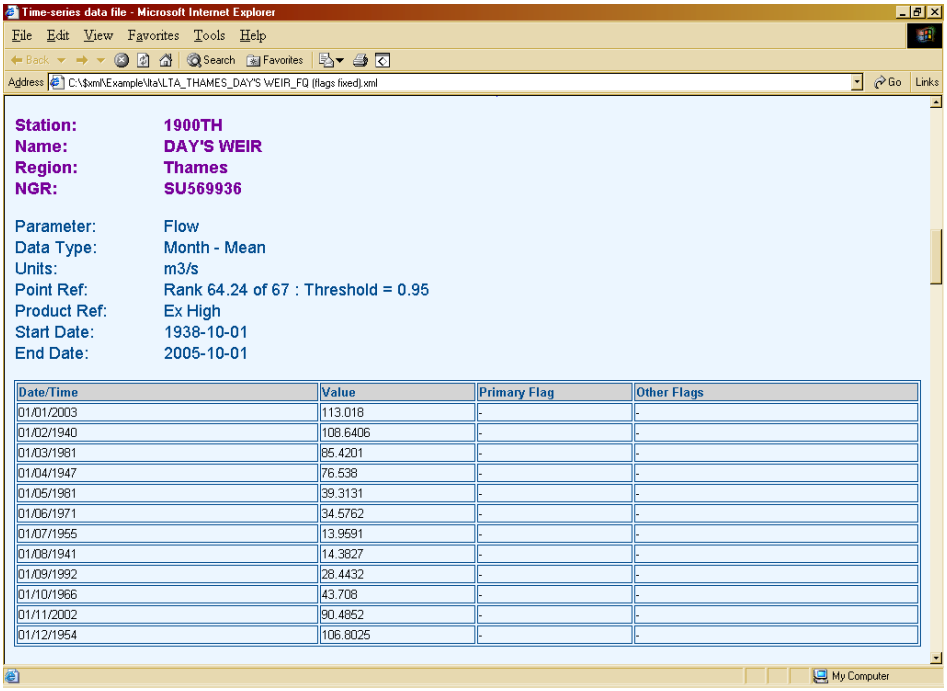
4. Click on the "Mark up using stylesheet" button, as indicated by the red arrow in *Figure 3*. You will be asked to find an XSLT stylesheet. You should find the "tsx-

[4] NB this must be a file that meets the EA's time-series data exchange standard.
[5] You should make sure that you are using version 4.2 or later.

Trans_LTA Probability Bands.xsl" stylesheet, which will perform your LTA probability ranking.

5. When it has completed the ranking process, it will generate a new XML time-series file that contains the results. You will need to give it an appropriate name (and location) to save it to.

And that is all there is too it!




| Date/Time | Value | Primary Flag | Other Flags |
|---|---|---|---|
| 01/01/2003 | 113.018 | - | - |
| 01/02/1940 | 108.6406 | - | - |
| 01/03/1981 | 85.4201 | - | - |
| 01/04/1947 | 76.538 | - | - |
| 01/05/1981 | 39.3131 | - | - |
| 01/06/1971 | 34.5762 | - | - |
| 01/07/1955 | 13.9591 | - | - |
| 01/08/1941 | 14.3827 | - | - |
| 01/09/1992 | 28.4432 | - | - |
| 01/10/1966 | 43.708 | - | - |
| 01/11/2002 | 90.4852 | - | - |
| 01/12/1954 | 106.8025 | - | - |

***Figure 4 – The Extremely High rank, which has been created by the "tsx-Trans_LTA Probability Bands.xsl" stylesheet***

See *Figure 4* as an example[6] of the results. Note that you are looking at the "Extremely high" rank. The actual results are interpolated so that they do meet the probability threshold (0.95 in this case) this information can be read from the *Point Reference*. In the set-of-values you will see that values are date-stamped, the date is taken from the more extreme value (in this case, that of rank 65).

Your source XML datasets will have been ranked and new time-series generated for:

- Maximum
- Extremely High
- Notably High
- Above Normal
- Below Normal
- Notably Low

[6] Notice that the results have been marked-up using the "tsx-Trans_EA Standard XHTML (not sorted).xsl" stylesheet so that it is easier to read them – see Appendix 4 for information on how to apply these formatting stylesheets.

- Extremely Low
- Minimum

These will all be output in XML time-series format so that they can be manipulated further. All of the other information (like: station name, reference, parameter, units, etc.) are transferred into the result file as a copy from the original file.

---

## Procedure 2 – Creating a CSV[7] table of your results

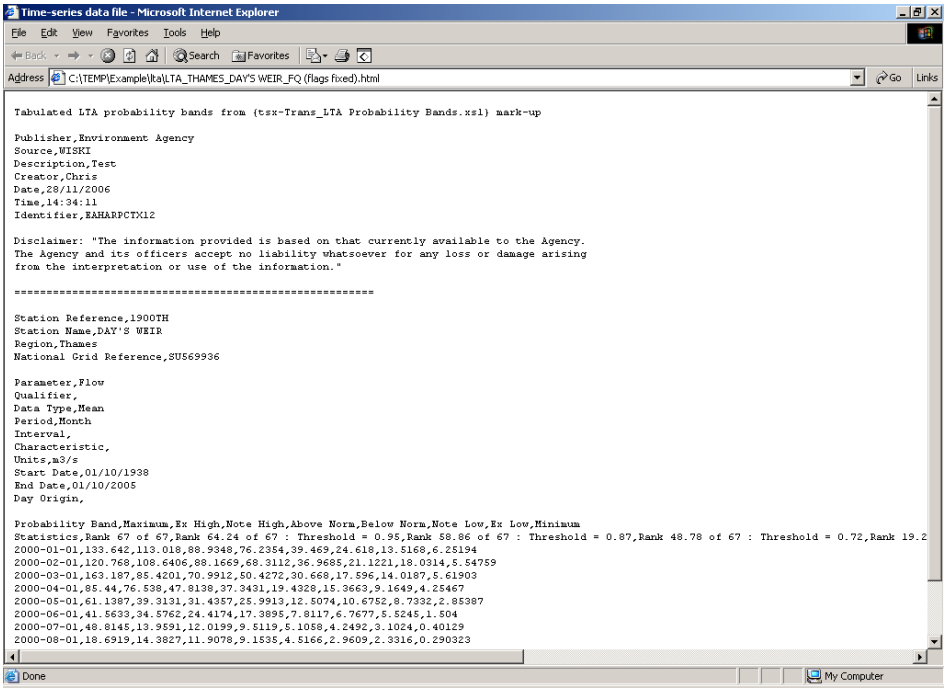



***Figure 5 – Results of "tsx-Trans_Table of Probability Bands (CSV).xsl" mark-up[8]***

It is likely that you want to put your results into a spreadsheet. In which case you will probably find it useful to use the "tsx-Trans_Table of Probability Bands (CSV).xsl" stylesheet to convert your XML results into CSV format (e.g. see *Figure 5*).

The procedure to create the CSV table of the ranks is the same as you used to generate those ranks from the original XML file.

1. Similar to instruction 3 – find the XML output file that you created from the ranking stylesheet.

2. Click on the "Mark up using stylesheet" button, in the Analyser, and then find the "tsx-Trans_Table of Probability Bands (CSV).xsl" stylesheet, which will format your data into a table.

---

[7] CSV = Comma Separated Values, a popular text-file format for spreadsheets, etc.
[8] Notice that this is being viewed through a web browser, the results could also be viewed in a text editor or read directly into a spreadsheet (like Excel).

3. When it has completed formatting, you will need to save the output to an appropriate filename.

You can either save it as:

- A webpage – like in *Figure 5* – by giving it a ".html" file extension (e.g. c:\ResultTable.html). Note that you can copy-and-paste the text from here.

- A CSV text file, by giving it a ".csv" file extension (e.g. c:\ResultTable.csv). This will load easily into a spreadsheet but you will need to ignore the top 9 lines and the bottom 3 lines.

---

## Procedure 3 – Building a graph

The design of this Table of Probability Bands (CSV) output has been written to be similar to the "tsx-Trans_EA Standard CSV text.xsl" stylesheet output. If you wish to create a graph, comparing a specific period of data against the LTA bands at the site, you may wish to use the "tsx_Monthly Data comparison v1'1.xls" spreadsheet.

These are high-level instructions (they assume that you are quite comfortable with Excel).

1. Mark-up your XML data, using the same procedure as described in the previous section, with the "tsx-Trans_EA Standard CSV text.xsl" stylesheet. Save it as a ".csv" file.

2. Open this in Excel and Copy the sheet.

3. Open the "tsx_Monthly Data comparison v1'1.xls" spreadsheet. Go to the "DATA INPUT" page and select all. Then Paste your data into the sheet (NB make sure you are on cell "A1" before you do).

4. Open the LTA probability bands CSV file that you created previously. Again, Copy this sheet.

5. Go back to the "tsx_Monthly Data comparison v1'1.xls" spreadsheet. Go to the "RANK BANDS INPUT" page and select all. Then Paste your data into the sheet (NB make sure you are on cell "A1" before you do).

6. Technically the "Table" and "Graph" should now be displaying a comparison of your input data, against the LTA probability ranks (e.g. see *Figure 1*).

You may need to tidy some things up at this point, especially if you have a very long record.

- The table is generated for down to row 1002, you may need to copy down further.

- You will probably need to alter the axes on the graph.
- You may also want to extend the range of the source data for the graph (note that this is only down to row 102 at the moment).
- You may also wish to tinker with labels, etc.

## Appendix 1 – Doing a whole batch of sites easily

If you want to carry out the LTA ranking process on a number of sites, the Analyser tool can help you to do this easily.

- Make sure all of the XML files that you want to process are located together, in the same directory.
- In the Analyser, find one of those files (does not matter which), and tick the "Whole Directory?" box.
- When you do the "Mark up using stylesheet" operation[9], it will work through all of the XML time-series files in your source directory.

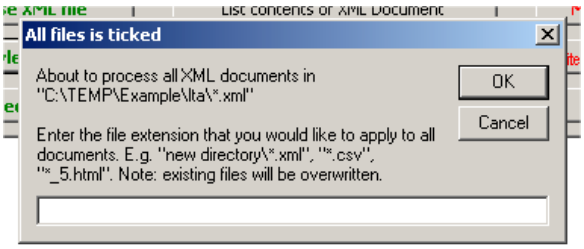



***Figure 6 – Marking-up a whole directory of XML files***

- It will request a file extension from you (see *Figure 6*), which enables it to automatically generate output files to store the results. Notice that you can include directory paths, and extensions to the file name (e.g. "LTA_*.xml"), to keep your results distinguishable from your input.

I advise setting up something like the following directories before you start: "`\raw`", "`\lta_bands`" and "`\lta_csv`".

1. Put all of your source XML files, that you want to rank, into the "raw" directory.

2. When you create the LTA bands: find one of those files in the "raw" directory; tick "Whole directory?"; and when the input box (*Figure 6*) pops up, type "`..\lta_bands\lta_*.xml`".

3. Your results will all be written to the "lta_bands" directory, and will all have "lta_" at the start of their names (e.g. from "file1.xml" the result will be "lta_file1.xml").

4. To then create CSV tables, remember to find one of the files in the "lta_bands" directory, before marking-up using the "tsx-Trans_Table of Probability Bands (CSV).xsl" stylesheet. You will then want to output to: "`..\lta_csv\*.csv`".

If you want to generate graphs of each site, you can mark-up all of your comparison data files against the "tsx-Trans_EA Standard CSV text.xsl" stylesheet, in one batch. However, generating the graphs is currently a manual process.

[9] This will work for any of the buttons with Green writing on them.

## Appendix 2 – Dealing with missing data

It is important to note that the "tsx-Trans_LTA Probability Bands.xsl" stylesheet does not do anything clever to handle missing data. This is deliberate because holes in your dataset could skew the statistics that you are generating. There is no obvious way to automatically fill gaps in data. The best approach is to fix them first.

- It is best to use proper infilling techniques, and fix the gaps on the hydrometric/hydrological archive before exporting the data.

- Otherwise, you should at least plot the data up so that you can judge whether the missing data is likely to come from a period of high, normal or low values. Having worked that out, you can make on the spot infills to try and make sure you are not skewing your results. Note that you can use the XML Analyser to help with this:

  1. Find the XML file that you want to edit.

  2. Click on the red "Mark up into a new spreadsheet" button (see *Figure 3*). NB this is not designed for large files so you will have to be patient.

  3. Close the Analyser (click on the [x] in the top-right of the window is easiest).

  4. Edit the spreadsheet that you have just created, being careful to maintain the format of the data.

  5. Re-open the Analyser, and click on the red "Write XML (source same as spreadsheets mark-up)" button.

  6. This will create a new XML file, which you can save and then process.

It is often easy to miss the presence of missing data in a set-of-values. Note that the LTA ranking process will bring these out. You can see in *Figure 7* (note that I have highlighted them in yellow), values of NaN – which mean Not-a-Number – appear in the minimum rank. If you see this, you should go back to your source data; fill the gaps; create a new, complete XML file; and re-run the process.

| Probability Band | Maximum | Ex High | Note High | Above Norm | Below Norm | Note Low | Ex Low | Minimum |
|---|---|---|---|---|---|---|---|---|
| Statistics | Rank 67 of 67 | Rank 64.24 of 67 : Threshold = 0.95 | Rank 58.86 of 67 : Threshold = 0.87 | Rank 48.78 of 67 : Threshold = 0.72 | Rank 19.22 of 67 : Threshold = 0.28 | Rank 9.14 of 67 : Threshold = 0.13 | Rank 3.76 of 67 : Threshold = 0.05 | Rank 1 of 67 |
| January | 133.642 | 113.018 | 88.9348 | 76.2354 | 39.469 | 24.618 | 13.5168 | 6.25194 |
| February | 120.768 | 108.6406 | 88.1669 | 68.3112 | 36.9685 | 21.1221 | 18.0314 | 5.54759 |
| March | 163.187 | 85.4201 | 70.9912 | 50.4272 | 30.668 | 17.596 | 14.0187 | 5.61903 |
| April | 85.44 | 76.538 | 47.8138 | 37.3431 | 19.4328 | 15.3663 | 9.1649 | 4.25467 |
| May | 61.1387 | 39.3131 | 31.4357 | 25.9913 | 12.5074 | 10.6752 | 8.7332 | 2.85387 |
| June | 41.5633 | 34.5762 | 24.4174 | 17.3895 | 7.8117 | 6.7677 | 5.5245 | NaN |
| July | 48.8145 | 13.9591 | 12.0199 | 9.5119 | 5.1058 | 4.2492 | 3.1024 | NaN |
| August | 18.6919 | 14.3827 | 11.9078 | 9.1535 | 4.5166 | 2.9609 | 2.3316 | NaN |
| September | 38.6367 | 28.4432 | 12.9081 | 9.2852 | 4.544 | 3.1037 | 2.6222 | 1.73967 |
| October | 74.5742 | 43.708 | 30.1183 | 18.9803 | 5.413 | 3.9704 | 3.4892 | 2.78226 |
| November | 128.13 | 90.4852 | 68.0332 | 45.036 | 10.3251 | 5.7313 | 4.4225 | 3.751 |
| December | 128.71 | 106.8025 | 70.4404 | 58.5954 | 27.3468 | 11.6131 | 7.1895 | NaN |

***Figure 7 – Example of missing data in your results***

# Appendix 3 – Exporting XML time-series from Wiski (the current EA Hydrometric Archive)

*You may wish to use the Wiski Export Wizard instead of options 1 to 3.*

1. Select the time-series that you want to export from Wiski Explorer. Note that you may want to set up a "Group Type" to make this efficient[10]. You can hold down the [control] or [shift] key to multi-select. See *Figure 8*.

2. Set the start and end time range at the top of the Explorer window.

3. Right-click on one of your selected time-series, and then select "Export" from the pop-up menu.

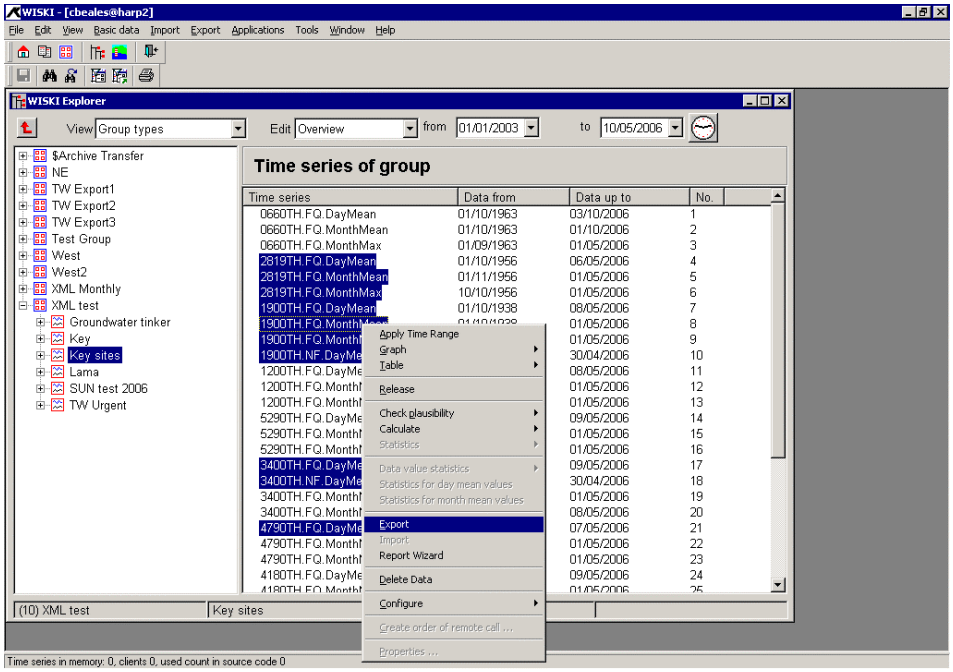



***Figure 8 – Selecting time-series for export in Wiski Explorer***

4. The data export window will open (see *Figure 9*).

5. Change directory to a convenient location on your hard disk (i.e. the V: drive).

6. Select the "EA XML Format" from the options in the list on the left.

[10] For example there is a "Key Flow Sites" group under the "Hydrology" group type that can be used by Regional Hydrology.

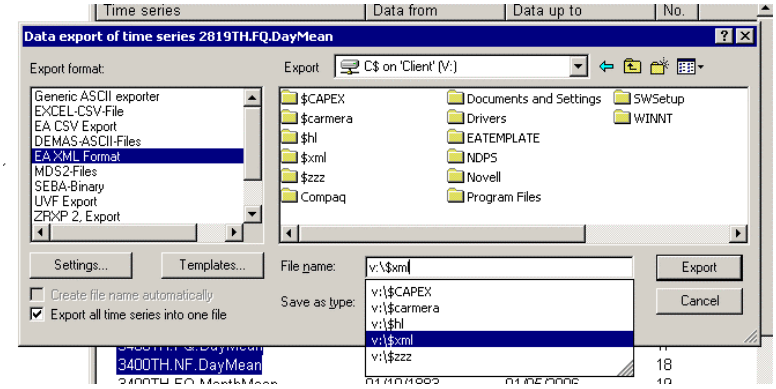



***Figure 9 – Export window***

7. Click the "Settings…" button and the settings window will pop up (see *Figure 10*).

8. Check that you are outputting "Saved Value" and the other settings look right (note that they should default ok). Note that if you are doing multiple exports for LTA analysis, make sure that the "File name" specification[11] is set to something like "`%w_%s_%p%i`" to make sure it automatically creates a filename. Click Ok when you are happy.

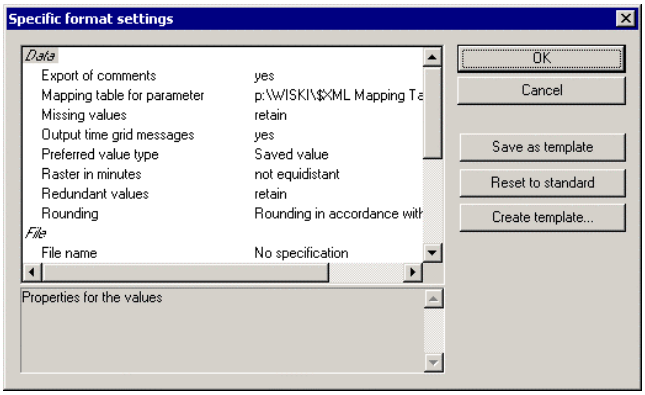



***Figure 10 – XML settings window***

9. Back at the Export Window, tick the "Create file name automatically" box, if you are wanting to create separate files for LTA analysis.

10. Alternatively, assuming you are doing multiple exports, you can click the "Export all time series into one file" box. If so alter the filename for export (make sure that it ends with ".xml").

11. Click the Export button.

12. When the time-series have been exported the file will be validated. Check to make sure that it has successfully created a valid file. If there are any errors, open the log file and potentially get Hydrometry to fix any issues with sites or data.

[11] These are all magic wildcards: %w = River, %s = Station Name, %p = Parameter (e.g. FQ = Flow, WL = Water Level), %i = Interval (e.g. 15 = 15 min).

# Appendix 4 – Using stylesheets to make it easier to read your XML data files

This document has explained how to use stylesheets to perform calculations on your data. It has then explained how to apply a "page-formatting" stylesheet, to create a CSV table of data, in a new file. However, you may wish to use some of the "page-formatting" stylesheets to look at your data on the fly.

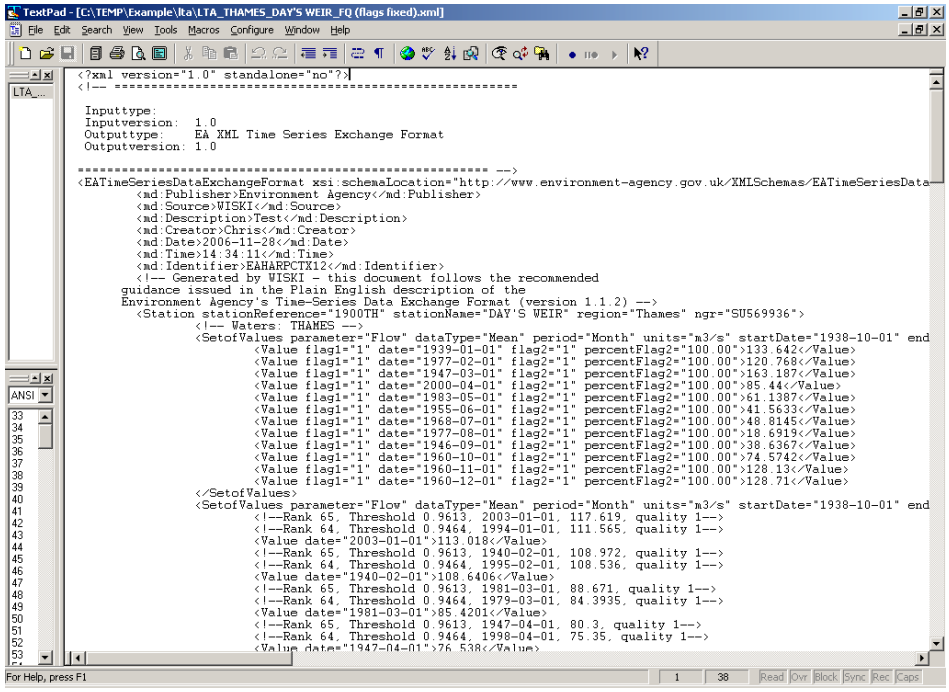



***Figure 11 – XML time-series data in a text editor***

XML data is all just text, as you can see from *Figure 11*, and it is "human readable[12]". However, many people will look at it and think it is a bit of a bonkers mess…lots of <brackets> and things…

You can also open up your XML files in a web browser like Internet Explorer, see *Figure 12*. Notice that it is easier to read the information in here, as Internet Explorer paints it with colours, and it also allows you to collapse and expand blocks of the data (look for the "+" and "–" characters on the top-left of each block).

But there is a better way…

[12] One of the objectives of the World Wide Web Consortium, www.w3c.org

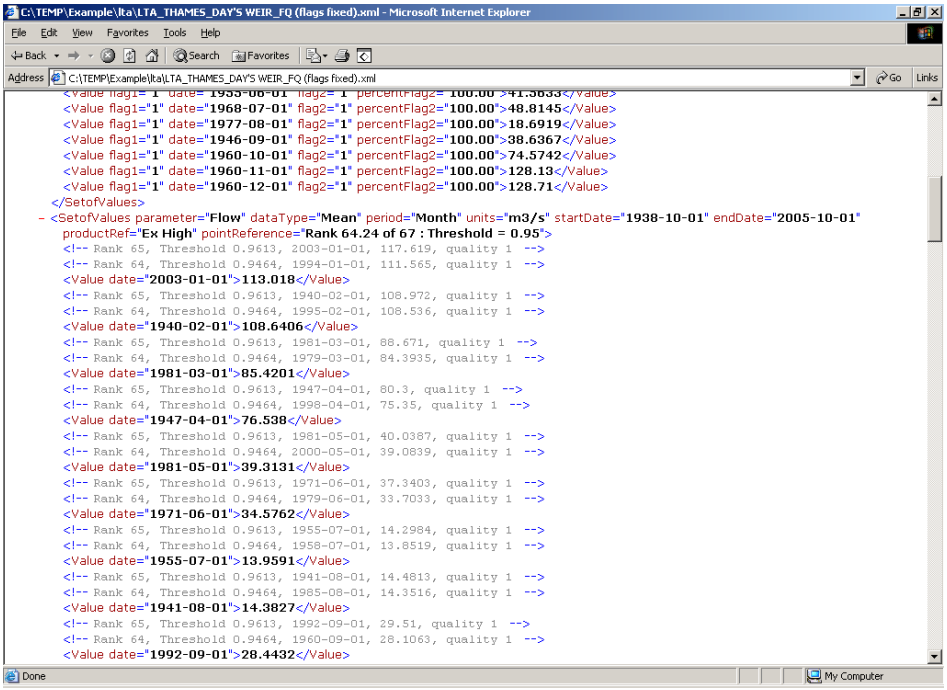



***Figure 12 – The same XML file as Figure 11, opened in Internet Explorer***

…If you get your XML file marked-up using the "tsx-Trans_EA Standard XHTML.xsl" stylesheet, it will look like the views in *Figure 4*, *Figure 13* and *Figure 14*. Note that these screenshots are taken from different parts of the file (scrolling up and down).

The "tsx-Trans_EA Standard XHTML.xsl" stylesheet, as you can see, marks-up the XML file into a webpage (i.e. XHTML[13]). This has a number of useful features (besides just making the information easier to read because it is not mixed up in XML tags):

- Data is formatted into tables
- A table of contents is created, which is helpful for navigating between sets-of-values
- Validated data, of different qualities, are coloured to help them stand out
- Data comments are shown next to the values in the data table
- Links to reference material are also available

[13] XHTML = eXtensible Hyper-Text Mark-up Language. It is the new version of HTML, which is the language used to write webpages.

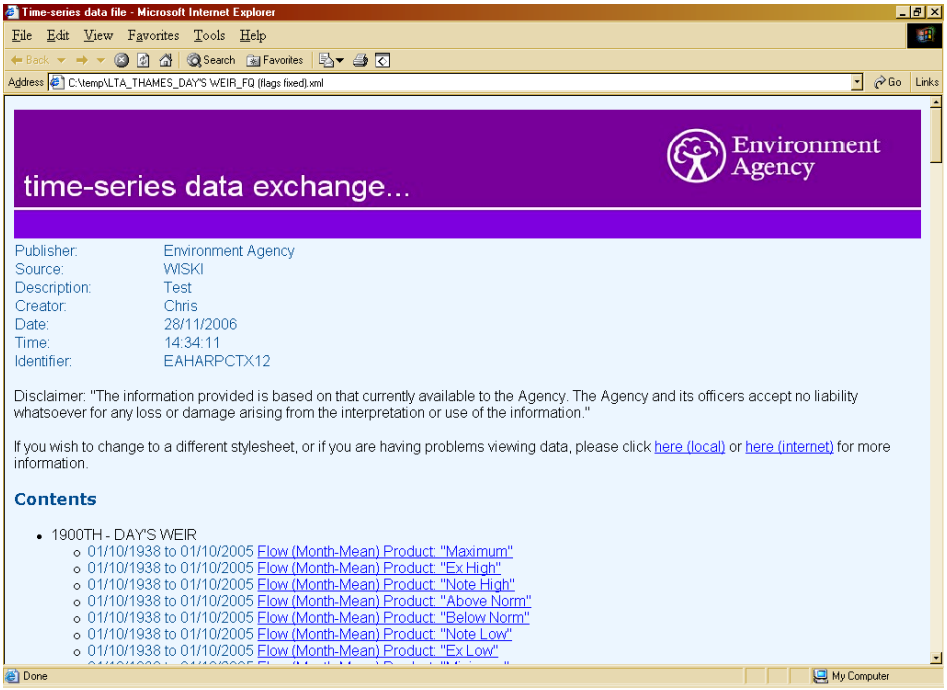



***Figure 13 – Same file as Figure 11 and Figure 12, marked-up using "tsx-Trans_EA Standard XHTML (not sorted)[14].xsl". Note that if you scroll down from here (or click on the link to* Flow (Month-Mean) Product: "Ex High"*) you will see the Extremely High data as in Figure 4***

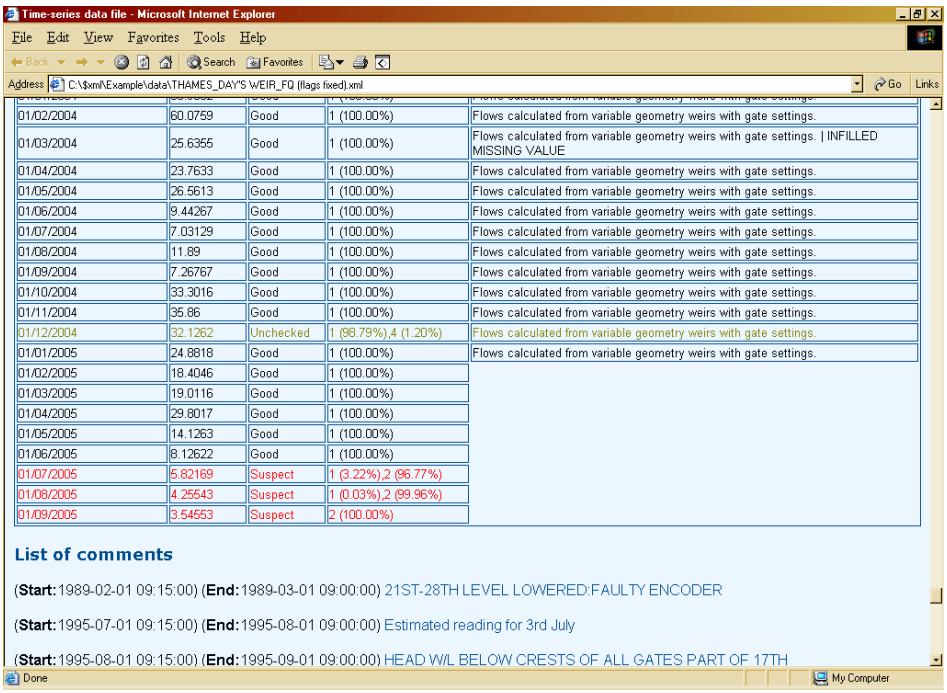



***Figure 14 – Part of another XML time-series file (this is actually the source data we used to generate LTA bands for the above Figures). This is showing the bottom of the data table. Note the colours for different data qualities, and that the comments are mapped against the relevant data values.***

[14] The "…not sorted…" stylesheet is exactly the same as the EA Standard XHTML, it is just a version which does not sort the values in each table (there is a switch inside, which is set to False).

Like with any stylesheet that is written for XML time-series files (like the ones we have used in generating the LTA probability bands) you can get the XML Analyser to mark-up your XML file and create a new file, with the results. If you use the "tsx-Trans_EA Standard XHTML.xsl" stylesheet, you would be creating a new ".html" webpage.

However, there is a better way of doing this. This involves adding a reference, to the stylesheet, and putting it in your XML file. The XML Analyser can make it very easy to do this (see *Figure 3*), all you need to is find your XML file, and then use the green "Add stylesheet reference" button[15].

- This will also work with the "Whole Directory?" box being ticked – in which case it will work through all the XML time-series files, in the same directory as the XML file you found at the top.

With a stylesheet reference added to an XML file, the XML data inside it remains unchanged. This is important because it means you can use all of the processing tools, other stylesheets, etc, on your data. However, whenever you open it in a web browser, the browser will mark it up, on the spot, so you can see the information in a nicely formatted way that really is easy to read.

- The only warning to give about stylesheet references is that you must remember that the reference points to the stylesheet and the directory it is in. Therefore, if you move your XML file to a different location, the web browser may not be able to find the stylesheet. You can always use the Analyser to change the reference (NB the "Add stylesheet reference" button can be used to change references as well). Alternatively, you might want to put your page-formatting stylesheets in the same directory as your XML files, and to move them together if you ever want to relocate them.

---

I hope this all makes sense. There is a lot more fun to be had with stylesheets so watch this space. :o)

Chris Beales
November 2006

[15] There is a "Remove stylesheet ref" button as well, which can be useful if you find it useful to view the raw XML in Internet Explorer.