Automated feature recognition in CFPD analyses of DMA or supply area flow data

APPENDIX 1: THE CFPD METHOD

In this section we give a summary of the description presented in Van Thienen (2013). Consider a supply area for which the flow rate into the area (accounting for all inflow, outflow and storage) is registered for a period of time (e.g. a day, a week, a month or an entire year) and again for a comparable period of the same length in another year. The registered patterns are likely to be similar in shape but not exactly the same. The simple CFPD procedure allows a quantitative comparison of these patterns, taking the following steps:

1. Sort both data sets from small to large magnitude. Sorted measurement ranks, scaled to a 0–1 range, are on the horizontal axis, flow rates are on the vertical axis.
2. Plot one data set against the other in a CFPD plot.
3. Determine a linear best fit with slope a and intercept b.

Note that the word pattern is used here in the sense of a time series which is generally repetitive to a significant degree with some variations. In general, it is preferable to construct the CFPD plot with the first period on the horizontal axis and the second on the vertical. In this case $a > 1$ and/or $b > 0$ corresponds to an increase in flow rate. Note that comparison of periods of different length is also possible but requires an additional interpolation step, see Van Thienen (2013).

For the application of the CFPD procedure on long time series, it is desirable to perform a comparison of each period (which will be called *block* in the following) within this time series with each other period. This allows the identification of changes on the timescale of individual blocks.

Figure A1 illustrates the procedure and results of such a block analysis. A CFPD analysis is made (Figure A1(a)) of all possible combinations of time blocks of a preselected length of the comparison frame within the complete dataset. Two matrices A (Figure A1(b)) and B (Figure A1(c)) are made, in which row i and column j represent blocks i and j (within the time series), respectively, and entries A_{ij} and B_{ij} are the factors a and b, respectively, resulting

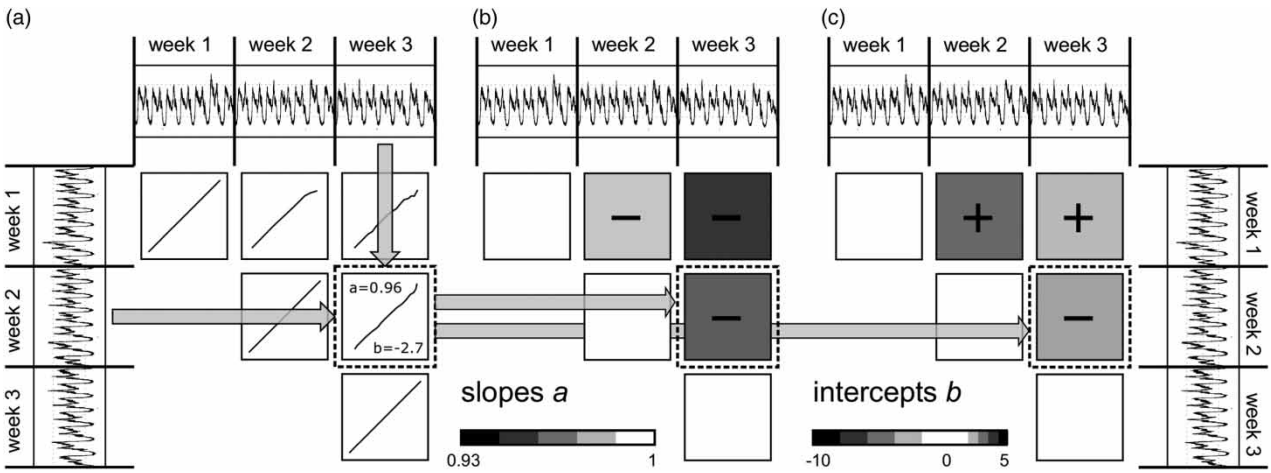

Figure A1 | Illustration of the CFPD block analysis. (a) CFPD analysis for each combination of blocks, (b) visualization of slope values (matrix A), (c) visualization of intercept values (matrix B). Copied from Van Thienen (2013).

from a CFPD comparison of block i with period j. The entries in the upper triangle (the lower triangle is not shown, as the matrices are antisymmetric) are grey toned or colored as a function of their deviation from 1 (A) and 0 (B), respectively, with small deviation having a light tone close to white and larger deviations having either a darker tone and a sign (−/=/+) indicating the direction of the deviation, or a red (+) or blue (−) color. The complete matrices are constructed because it is usually not clear beforehand which time block is suitable as a reference time block.

Changes in a or b which remain in the signal longer than the frame length will show up in the block analysis as blocks of similar tone and sign, allowing direct pinpointing (in time) of events which cause these changes.

REFERENCE

Van Thienen, P. 2013 A method for quantitative discrimination in flow pattern evolution of water distribution supply areas with interpretation in terms of demand and leakage. *J. Hydroinform.* **15** (1), 86–102.