A method to apply national and spatially variable scaling factors to abstraction data over England

Helen Baron, UKCEH

March 2025

1 Introduction

This script, `abstraction_scaling.py`, multiplies the baseline abstraction data for England (`England_Monthly_Abstractions_1km_GW_SW_TW_199901_201412.csv` available from [1]) by annual scaling factors to produce abstraction values for different future scenarios. It has been designed to produce future abstraction scenarios as part of the CS-N0W project [2], but is intended to be flexible so that new data and methods can be employed.

A secondary function of this script is to take the baseline abstractions in csv format and convert them to netcdf files, with abstractions grouped by water-use sector as described in section 5.1.

2 The scenarios

This set-up has three scenarios, as described in [2], but the method can be adapted to any number of scenarios provided the scaling factor file is updated accordingly. Within the script, the number of scenarios is set by `scen_nos` (see Table 4).

The scaling factors used to produce future abstractions for the CS-N0W project are provided in `scaling.csv` and the associated files (`PWS_scaling_Sus.csv`, `PWS_scaling_BaU.csv`, and `PWS_scaling_EG.csv`) and their derivation is described in the future abstractions dataset documentation [3].

The scaling factors are applied to a five year average of the baseline abstractions (1999 to 2014 inclusive) as described in [3], these averages are calculated from the baseline data within the script.

3 The scaling factors

3.1 Water-use sectors

The scaling factors vary for different water-use sectors, as listed in the **Name** column of the scaling factor file (this column is for reference and is not used in the script). The water-use sectors are specified by the **Primary Code**, **Secondary Code** and **Use Code** columns (these columns appear in the baseline abstraction data, and are explained in Appendix Tables A1-3 in [4]), and for each of the water-use sectors there are n rows to specify the groupings (where n=`scen_nos` for convenience). The values in these columns must match the values that appear in the baseline data, or be listed as 'ALL' if the user does not want to filter on that particular column.

In the example scaling file given (`scaling.csv`), the food&drink sector includes all abstractions where **Primary Code** is I, **Secondary Code** is either FAD, BRW or DAR; and **Use Code** is one of the values listed; but also includes any abstraction with **Use Code** equal to 460 or 470 ("Vegetable Washing" and "Water Bottling" respectively). The PWS sector (public water supply) includes abstractions where **Primary Code** is W; **Secondary Code** is either PWS or WAT (thus excluding the PRV and PWU codes which apply to private water undertakings); and **Use Code** is one of the values listed which does not include 470 - "Water Bottling" - since these abstractions are scaled according to the food&drink sector.

The sector scaling factors are only applied to the abstractions which match the given patterns. Any abstraction not covered by the scaling factors is kept constant in the future scenarios.

3.2 National and spatial scaling

The scaling factors are all annual, and are either national (i.e. one number for each year to be applied across the entire dataset) or spatial (i.e. vary geographically, with an annual value for each geographical unit). The national scaling factors are included in the scaling factors file, and the script reads these values and applies them to the abstractions for each water use sector (as described in section 3.1).

Spatial scaling factors must be stored in separate csv files, with the file name listed for each future scenario in the first year of the future projections. These files contain scaling factors for each year and each spatial unit, with the spatial unit identifier in the first column (this must be an integer).

A netcdf file is required to map these spatial units onto the baseline abstractions, with the file name listed under the **Spatial Map** column in the scaling file. This file must be at the same grid resolution as the baseline abstractions and have dimensions named Northing and Easting, and a variable whose name matches that of the spatial unit identifier column in the spatial scaling csv files. The script uses this netcdf file to allocate spatial identifiers to the baseline abstractions, and scale them according to the factors held in the spatial scaling factor files. Any abstraction that does not have a matching spatial identifier is kept constant in the future scenarios.

For the CS-N0W project, most of the scaling factors are applied nationally, except PWS which is applied by water resource zone (WRZ). The WRZ map provided is `EW_WRZ.nc` - a gridded version of England's WRZ shapefiles (i.e. not including areas in England which are supplied by Welsh Water). In the example given, the spatial scaling files are `PWS_scaling_Sus.csv`, `PWS_scaling_BaU.csv`, and `PWS_scaling_EG.csv`, and the spatial identifier is `WRZ_ID`.

4 Settings

At the start of the script are a set of variables and filepaths that need to be updated by the user, as described in Table 4.

Name	Description
`scen_nos`	The number of future scenarios. This should also match the number of rows allowed for each water-use sector in the scaling and grouping csv files.
`first_year`	The first available year for the scaling factors. This also corresponds to the column where the spatial scaling file names are listed if the scaling factors are spatial (see section 3.2).
`start_year`	The first year that you want to produce future abstraction data for.
`end_year`	The last year that you want to produce future abstraction data for.
`netcdf_output`	This must be `True` or `False`, depending if netcdf output files are required (see section 5.1).
`abstractions_filepath`	Full filepath of the baseline abstraction data (csv file).
`spatial_grid_filepath`	Full filepath of the spatial mapping data (netcdf file) - required if any of the scalings are spatial scalings.
`scaling_filepath`	Full filepath of the scaling factor data (csv file).
`grouping_filepath`	Full filepath of the grouping file (csv file) - required if netcdf outputs are desired (see section 5.1).
`savepath`	Path to the folder where the output files are to be saved.

Table 1: Table of user inputs within the abstraction_scaling.py script.

5 Output

The script outputs csv files in a similar format to the baseline abstraction data, but with the column headers in place, separate years in separate files, and with the addition of any spatial identifiers used for spatial scaling (see section 3.2). The naming convention for these files is:

`abstractions_<year>_<scenario>.csv`

for the years specified in the script, and the future scenarios detailed in the scaling factor file.

5.1 Grouping for netcdf

If required, the script can also output netcdf files of the abstractions grouped into different water-use sectors and split by source: groundwater, GW, and surface water, SW. This requires `netcdf_output` to be set to `True` in the script, and a grouping file in csv format (see the example file given: `grouping.csv`). The grouping file is similar to the scaling factor file: it has **Primary Code**, **Secondary Code** and **Use Code** columns which specify the abstractions which must be grouped together for a particular water-use sector, with n rows to specify the groupings (n=`scen_nos` for convenience). It has an additional column, **Source Code**, which can be used to group abstractions (see [**?**] for options). For each water-use sector, any abstractions which match the given patterns are included in a variable in the netcdf file, with variable attributes `standard_name` as given in the **Name** column, and `units` as given in the **Units** column.

The grouping file also contains a filepath for an existing netcdf file which is used to set the extent and resolution of the output netcdf files, as well as the a mask for non-valid values (i.e. a land-sea mask). The filepath must be listed in the first row of the grouping file under the **Mask** column, it must contain a variable named `mask`, and have the same resolution as the baseline data.

The naming convention for these files is:

`abstractions_<source>_<year>_<scenario>.nc`

A matching netcdf file for the baseline abstractions can be created if required, by running lines 249-250 in the script (currently commented out).

6 Running the script

To run the script, first ensure that your python set-up is suitable:

The required python packages are listed in `requirements.txt`, this can be used to create a virtual environment to run the script from. Alternatively, it runs successfully from the Jasmin Jaspy environment version: `jaspy/3.11/v20240815`.

Secondly, ensure the user-input settings are correct (see section 4).

Finally, run the script using: `python abstraction_scaling.py`

References

[1] P. Rameshwaran, V. Bell, H. Davies, P. Sadler, A. Beverton, R. Thornton, and M. Rhodes-Smith, "Gridded Actual Abstraction, Discharge and Hands-off Flow Datasets for England.," 2024.

[2] H. Baron, V. Keller, J. Hannaford, and V. Bell, "Approaches to construct scenarios of future water demand – D2: Future water availability for water intensive energy infrastructure. Climate Services for a Net Zero Resilient World (CS-N0W) report," June 2023.

[3] H. Baron, V. Keller, P. Rameshwaran, A. Beverton, J. Wilson, V. A. Bell, H. N. Davies, and J. Hannaford, "Data documentation: Projected Abstractions and Discharges for England, 2020 to 2080, for three future scenarios," 2025.

[4] P. Rameshwaran, V. Bell, H. Davies, P. Sadler, A. Beverton, and R. Thornton, "Data documentation: Gridded actual abstraction, discharge and hands-off flow datasets for england.," 2024.