

README File for 'MPETUS Water Demand Microsimulation Model output'

Data DOI: <https://doi.org/10.5258/SOTON/D1815>

Ben Anderson

Last run at: 2021-05-19 12:14:57

Contents

1 Background.....	2
2 License.....	2
3 The data	2
3.1 Contextual variables	4
3.2 Model variables	4
3.3 Estimated daily water demand variables (l/day).....	4
4 Basic plots.....	5
5 All data (descriptives).....	7
6 End of file	9

1 Background

This file describes the output data from the IMPETUS domestic water demand microsimulation models.

The permanent data DOI is:

- <https://doi.org/10.5258/SOTON/D1815>

This dataset supports a number of publications which can be found at:

- <https://eprints.soton.ac.uk/cgi/search/simple?q=IMPETUS>

The code used to build the models is available from:

- <https://git.soton.ac.uk/SERG/IMPETUS/-/tree/master/impetusModel>

The original data used to construct the models is generally not publicly available but comprised:

- the SPRG/ARCC Domestic Water Use Survey - see Browne, Alison, Ben Anderson, and Will Medd. 2011. 'Lancaster SPRG/ARCC Water Survey'. Project Report. <https://eprints.soton.ac.uk/439637/>.
- seasonal demand adjustment coefficients derived from Parker, Joanne M. 2014. "Assessing the Sensitivity of Historic Micro-Component Household Water-Use to Climatic Drivers." A Doctoral Thesis. submitted in partial fulfilment of the requirements for the award of Doctor of Philosophy of Loughborough University, Loughborough: Loughborough. <https://dspace.lboro.ac.uk/2134/14939>.
- CEH reconstructed drought histories for the Colne catchment in East Anglia for 1994-2012 from Parry, Simon, Robert L. Wilby, Christel Prudhomme, and Paul J. Wood. 2016. 'A Systematic Assessment of Drought Termination in the United Kingdom'. Hydrology and Earth System Sciences 20 (10): 4265.
- MetOffice monthly rainfall and temperature data for the East of England for 1994-2012 sourced from <https://www.metoffice.gov.uk/research/climate/maps-and-data/uk-and-regional-series>

2 License

- CC-BY <https://creativecommons.org/licenses/by/4.0/>

3 The data

The version of the data file to which this README refers was created at 2021-05-18 17:16:09 and is approximately 44 MB in size (gzipped).

The data file contains a record for every SPRG household (n = 1633) projected forwards for each month in the period 1994 - 2012 (with incomplete years in 1994 and 2012) for each of a number of model versions.

This means that the number of records per model per year are as shown in Table 3.1 and the total number of records is 705,456.

```
t <- dt[, .(nRecords = .N), keyby = .(year = lubridate::year(obsDate), modelDef)]

kableExtra::kable(dcast(t, year ~ modelDef), caption = "N records per model per year")
%>%
  kable_styling()
## Using 'nRecords' as value column. Use 'value.var' to override
```

Table 3.1: N records per model per year

year	Baseline WE	Enhanced WE
1994	1633	1633
1995	19596	19596
1996	19596	19596
1997	19596	19596
1998	19596	19596
1999	19596	19596
2000	19596	19596
2001	19596	19596
2002	19596	19596
2003	19596	19596
2004	19596	19596
2005	19596	19596
2006	19596	19596
2007	19596	19596
2008	19596	19596
2009	19596	19596
2010	19596	19596
2011	19596	19596

Table 3.1: N records per model per year

year	Baseline WE	Enhanced WE
2012	17963	17963

The variables included are described below and broadly represent contextual or meta data together with estimated litres/day per household for various water uses under various models and scenarios.

3.1 Contextual variables

Limited contextual data that can be released:

- ID - a hashed version of the original SPRG survey household id (to eliminate disclosure risk)
- model - the version of the model used to estimate this data point (ref code/papers above for explanation)
- modelDef - model version as text (ref code/papers above for explanation)
- obsDate - the mid-point of the month for this data point
- currSeason - the season of the month for this data point
- metered - whether the household is metered
- occRed - count of occupants according to the SPRG survey (1-6+)
- Colne - the seasonal drought history for the season/year of the data point

3.2 Model variables

- loWC, loSh, dualFlushWC, loFlowShower, loSh - variables used to develop the microsimulation of water demand over time (ref code above for explanation)

3.3 Estimated daily water demand variables (l/day)

- Basin - bathroom basin use
- Bath - bath use
- Dishwasher - dishwasher use
- External - garden etc use
- KitchenSink - water use in the kitchen
- Shower - shower use
- WC - WC water use
- WashingMachine - washing machine water use
- SumDaily - sum of the above

Each of the above water uses appear in three forms:

- *.madj - monthly adjusted values before any water efficiency or drought interventions
 - *.madj.we - monthly adjusted values after water efficiency interventions are applied
 - *.madj.we.dr - monthly adjusted values after water efficiency and drought interventions based on the CEH drought history are applied

4 Basic plots

To confirm data distributions, [4.1](#) shows the mean of daily household water use totals by model definition.

```
madj.we <- dt[, .(meanSum = mean(sumDaily.baseline.madj.we, na.rm = TRUE)), keyby = .(obsDate, modelDef)]

madj.we$totalDef <- "Monthly adjusted with water efficiency"

madj.we.dr <- dt[, .(meanSum = mean(sumDaily.baseline.madj.we.dr, na.rm = TRUE)), keyby = .(obsDate, modelDef)]

madj.we.dr$totalDef <- "Monthly adjusted with water efficiency and drought"

plotDT <- rbind(madj.we, madj.we.dr)

ggplot2::ggplot(plotDT, aes(x = obsDate, y = meanSum, colour = totalDef)) +
  geom_line() +
  facet_grid(modelDef ~ .) +
  theme(legend.position = "bottom")
```

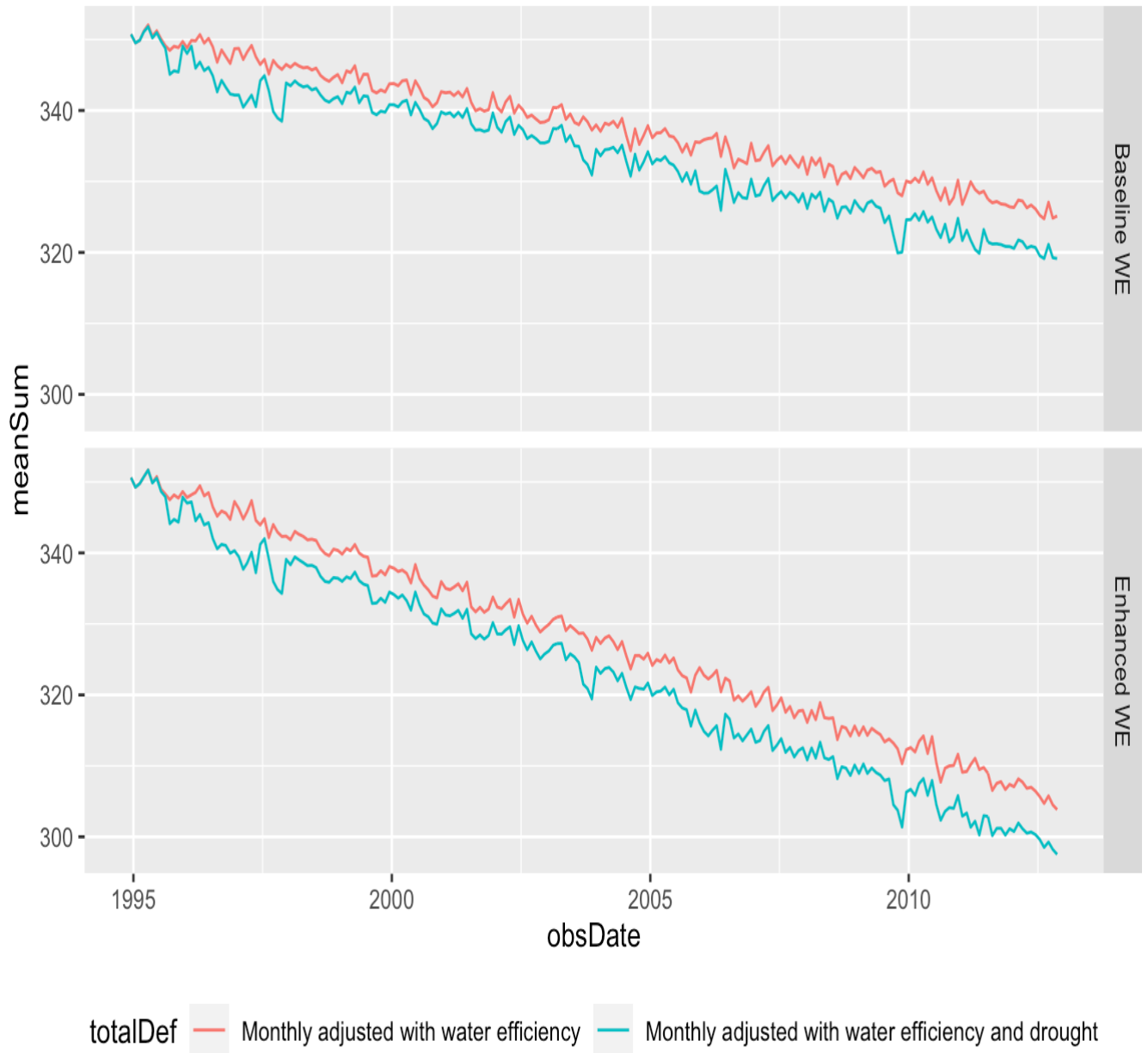


Figure 4.1: Mean of sumDaily per household by model definition

5 All data (descriptives)

The following table provides data descriptives for all variables included in the data package.

```
skimr::skim(dt)
```

Table 5.1: Data summary

Name	dt
Number of rows	705456
Number of columns	33
Key	NULL
<hr/>	
Column type frequency:	
character	10
Date	1
numeric	22
<hr/>	
Group variables	None

Variable type: character

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
model	0	1	4	4	0	2	0
metered	0	1	7	11	0	2	0
dualFlushWC	0	1	10	12	0	2	0
loFlowShower	0	1	8	11	0	2	0
currSeason	0	1	6	6	0	4	0
Colne	0	1	9	17	0	5	0
dualFlushWCdr	0	1	10	12	0	2	0
loFlowShowerDr	0	1	8	11	0	2	0
modelDef	0	1	11	11	0	2	0
ID	0	1	40	40	0	1633	0

Variable type: Date

skim_variable	n_missing	complete_rate	min	max	median	n_unique
obsDate	0	1	1994-12-15	2012-11-15	2003-11-30	216

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
Basin.baseline.madj	21600	0.97	57.24	52.73	10.33	26.56	39.88	66.28	463.65	█-----
Bath.baseline.madj	0	1.00	33.83	19.32	-3.95	19.82	26.12	53.57	77.53	█-----
Dishwasher.baseline.madj	0	1.00	1.78	3.16	-2.18	-0.65	-0.06	4.20	8.96	█-----
External.baseline.madj	72144	0.90	20.81	18.87	-1.28	1.06	19.37	32.99	98.17	█-----
KitchenSink.baseline.madj	0	1.00	19.91	10.15	7.47	15.12	15.43	30.26	76.12	█-----
Shower.baseline.madj	20736	0.97	105.85	76.28	-0.24	48.92	93.89	152.48	515.29	█-----
WC.baseline.madj	22896	0.97	88.27	47.35	20.64	49.50	74.54	112.39	262.79	█-----
WashingMachine.baseline.madj	15120	0.98	23.49	17.14	-3.24	14.71	17.83	30.27	83.44	█-----
sumDaily.baseline.madj	72144	0.90	350.54	182.45	53.53	209.02	312.18	475.51	1153.15	█-----
occRed	0	1.00	2.60	1.32	1.00	2.00	2.00	4.00	6.00	█-----
loWC	0	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	-----█
loSh	0	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	-----█
WC.baseline.madj.we	22896	0.97	74.43	46.71	9.81	36.21	70.52	100.08	262.79	█-----
Shower.baseline.madj.we	20736	0.97	102.13	75.04	-0.24	48.70	93.65	149.46	515.29	█-----
sumDaily.baseline.madj.we	72144	0.90	332.99	176.40	35.85	193.84	295.64	448.96	1153.15	█-----

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
dualFlushWCupdated	0	1.00	0.00	0.01	0.00	0.00	0.00	0.00	1.00	█-----
loFlowShowerUpdated	0	1.00	0.10	0.31	0.00	0.00	0.00	0.00	1.00	█-----
applyTUB	0	1.00	0.01	0.11	0.00	0.00	0.00	0.00	1.00	█-----
WC.baseline.madj.we.dr	22896	0.97	74.42	46.71	9.81	36.21	70.52	100.07	262.79	█-----
Shower.baseline.madj.we.dr	20736	0.97	98.15	73.48	-0.24	47.47	92.71	136.91	515.29	█-----
External.baseline.madj.we.dr	72144	0.90	20.25	18.58	-1.28	1.06	18.48	31.92	98.17	█-----
sumDaily.baseline.madj.we.dr	72144	0.90	328.52	174.30	35.85	191.52	291.45	442.10	1153.15	█-----

6 End of file