Review of Thames Water's Statement of Response to GARD's criticism of assessments of SESRO and STT deployable outputs in draft WRMP24

Contents

Summary

Refer to page no.

GARD's response to the consultation on Thames Water's draft WRMP24 was heavily critical of the assessments of the deployable outputs and drought resilience of Abingdon reservoir and the Severn to Thames transfer. This paper reviews Thames Water's dismissal of GARD's concerns which can be seen in Appendix G2 to their Statement of Response. My views on Thames Water's responses to GARD's concerns can be summarised as:

Regarding the validity of Pywr modelling and stochastic data

Thames Water has provided no valid argument to overturn GARD's view that the Pywr modelling of the London supply system is not fit for purpose. The problem with the Pywr modelling is in the validity of the stochastically generated river flows used in the modelling, rather than in the water supply system simulation in the Pywr model itself. There are two main problems:

- The stochastically generated Teddington natural flows are a poor fit to gauged natural flows and flows generated from historic weather data.
- Use of the period 1950-1997 to "train" the stochastic modelling has replicated the pattern of droughts in that period and excluded the long droughts that occurred in 1921, 1933-34 and 1943-44. 5-24

Regarding the deployable output of Abingdon reservoir

Thames Water has provided no valid arguments to overturn GARD's view that the DO of the 150 Mm^3 reservoir should be reduced from 271 Ml/d to 200 Ml/d, even before consideration of its lack of resilience to long duration droughts:

- Regarding double counting of droughts (6 Ml/d DO reduction), TW's argument that droughts spanning a year-end comprise two droughts is not valid. Nor is the 6 Ml/d error too small to be worth correcting. 26-28
- Neither is use of the incorrect value of the Culham MRF (2 Ml/d DO reduction) too small to be corrected – multiple small errors can add up to large errors. 28-31
- Regarding the use of the wrong climate change scenario (19 Ml/d DO reduction), in comparing Abingdon reservoir with other options, especially those resilient to climate change, TW should use the DO relating to the 'high' CC reduction as per their final WRMP. 31-34
- Regarding the allowances for dead and emergency storage (44 Ml/d DO reduction), TW's allowance of 30-day reservoir throughput for emergency storage is based on incorrect calculation of reservoir throughput for their existing reservoirs, which actually have emergency storage equivalent to about 70 days of throughput. The allowance of an average dead storage depth of only 2.5 m is yet to be justified by water quality modelling. 34-40

Refer to page no.

Regarding the resilience of Abingdon reservoir to long duration droughts

Thames Water has provided no valid evidence to overturn GARD's view that, if proper consideration is given to the occurrence of long duration droughts, the deployable output of the reservoir would be far less than that claimed by TW, perhaps in the region of only 50% of the claimed amounts:

- The method of method of generating stochastic flows has precluded the type of long drought that Abingdon reservoir cannot handle – three dry summers with two intervening dry winters 15-19 & 40-45
- TW have failed to follow WRSE's advice that, as extended droughts may not necessarily be well reflected in the stochastic data, drought artificial weather series should be used to represent prolonged drought events. 45-47

Regarding the deployable output of the unsupported Severn-Thames transfer

Thames Water has misunderstood GARD's point about their under-estimation of the DO of unsupported transfers by about 40% and have provided no evidence to overturn it:

- TW have provided no evidence of the validity of Pywr modelling of STT options or of the validity of River Severn stochastic flows used in the modelling.
- GARD's assessment of DOs using historic flows gives much higher DOs than TW's assessment of the 1:100 year DO which should be similar to the worst historic DO. 48-49

Regarding the need for United Utilities sources to replace Vyrnwy support for the STT

Thames Water has failed to address GARD's concern about over-estimation of the need for UU replacement sources for Vyrnwy support water, perhaps by about 70 Ml/d, which would have a substantial impact on the cost effectiveness of Vyrnwy support options. It is not acceptable for to TW to say that they have no knowledge of how UU calculated the need for replacement sources or to criticise GARD for not addressing their concern to UU. 49-52

Regarding estimates of utilisation of the STT

Thames Water has provided model output which shows STT utilisation for about 14% of the time (51 days per year). However, this is far less than the 22% utilisation said to have been assumed in the operating cost estimates shown in the Gate 2 report for the STT. 52-54

Ofwat's final Gate 2 decision reports on Abingdon reservoir and the STT

Ofwat's final Gate 2 reports have recognised GARD's criticisms of the DO assessments and included reasonable summaries of GARD's concerns. However, the only proposed actions were for the water companies to deal with GARD's concerns as part of the consultation response process. The evidence provided in this paper shows that this has not happened. 55-57

Therefore, these matters should now be assessed by a detailed and genuinely independent review by the regulators before any decisions are made on the final WRMPs.

1. Introduction

GARD's response to Thames Water's draft WRMP24 was heavily critical of the assessments of the deployable outputs and drought resilience for Abingdon reservoir and the Severn to Thames transfer. The criticism was detailed in the Addendum to the consultation response, which was subsequently replicated as an Addendum to GARD's response to Ofwat's Gate 2 decisions which can be seen here:

https://www.abingdonreservoir.org.uk/downloads/GARD%20response%20to%20Ofwat%20Gate%202,%20Volume%202%20Addendum%2011.5.23.pdf

GARD's criticisms were covered under three headings:

1. Validation of models and stochastic data – it was concluded that the modelling was not fit for purpose, primarily because of the use of invalid stochastically generated river flow data.

2. Deployable output of Abingdon reservoir – it was concluded that the true deployable output of the reservoir is about half the amount claimed by Thames Water and the reservoir would have little resilience to droughts longer than 18 month.

3. Severn to Thames transfer deployable output and operating costs – it was concluded that the deployable output of options with small amounts of support would be about 50% higher than the amounts claimed by Thames Water and that options with support from Vyrnwy reservoir had grossly over-estimated the need for replacement sources in the north-west. Usage of the transfer had also been over-estimated leading to over-estimates of operating cost and carbon impact.

Thames Water's Statement of Response has dismissed virtually all GARD's criticisms and there have been no significant changes in the revised version of WRMP24 submitted in September 2023.

This paper reviews Thames Water's dismissal of GARD's concerns which can be seen on pages 103 to 121 of Appendix G2 to their Statement of Response, as here:

https://dn9cxogfaqr3n.cloudfront.net/revised-draft/statement+of+response+on+draft+wrmp24/dWRMP24+SoR+Appendix+G2+-+Response+to+representations+from+organisations.pdf

2. Validation of models and stochastic data

2.1 Validity of the Pywr modelling and the hydrological model for generating river flows

GARD commented on the two step process used by Thames Water to validate their models:

1. Step 1 – validation of the modelling of existing London supplies (historical time series) using flow inputs taken directly from WARMS2. The aim of this validation step was to ascertain whether the Pywr simulation model replicates Thames Water's WARMS2 or Aquator simulation modelling, if using the same river flow data.

2. Step 2 - validation of the model (historical time series) using flow inputs derived by the new hydrological models which were then used for the stochastic modelling. The aim of this validation step was to ascertain the differences in model outputs caused by differences in river flows generated by the different hydrological models.

2.1.1 Step 1 – Validity of the Pywr simulation model

Regarding the Step 1, GARD commented that there was only a moderately good fit between Pywr and WARMS2 modelling of reservoir drawdowns, as shown by the plot below from Figure I-6 in the original Appendix I, and described by TW as "a very close match":

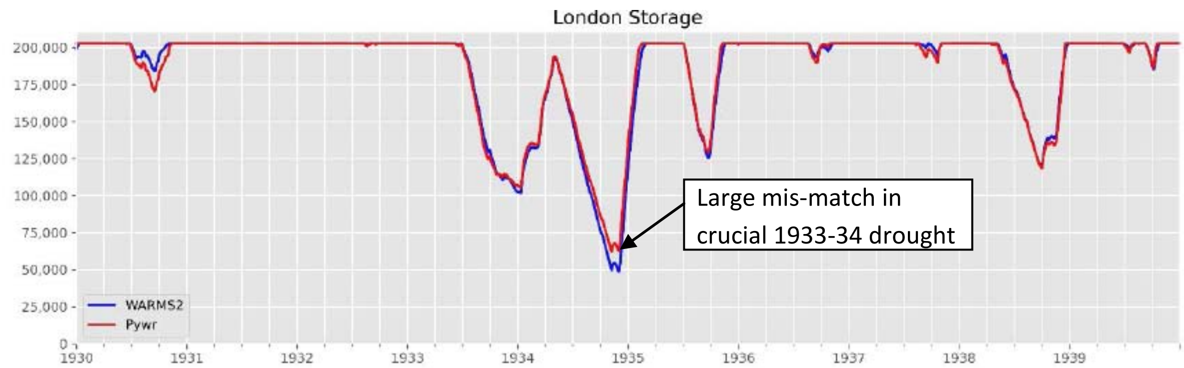


Figure 1 - TW Step 1 validation of Pywr, simulation of London storage using WARMS flows

TW have responded by saying that the plot above in their Appendix I was an out-of-date version and that more recent modelling, as per data supplied to GARD in March 2023, showed a better fit between the Pywr modelling and WARMS2 modelling when using the same river flow input data. I have checked this and agree that the revised version of the above plot shows a reasonably good match between the two models, as shown in Figure 2 below. The upper graph is from Figure I-6 in TW's revised Appendix I and the lower graph is plotted by me using the data provided by TW in file 'WARMS flow validation':

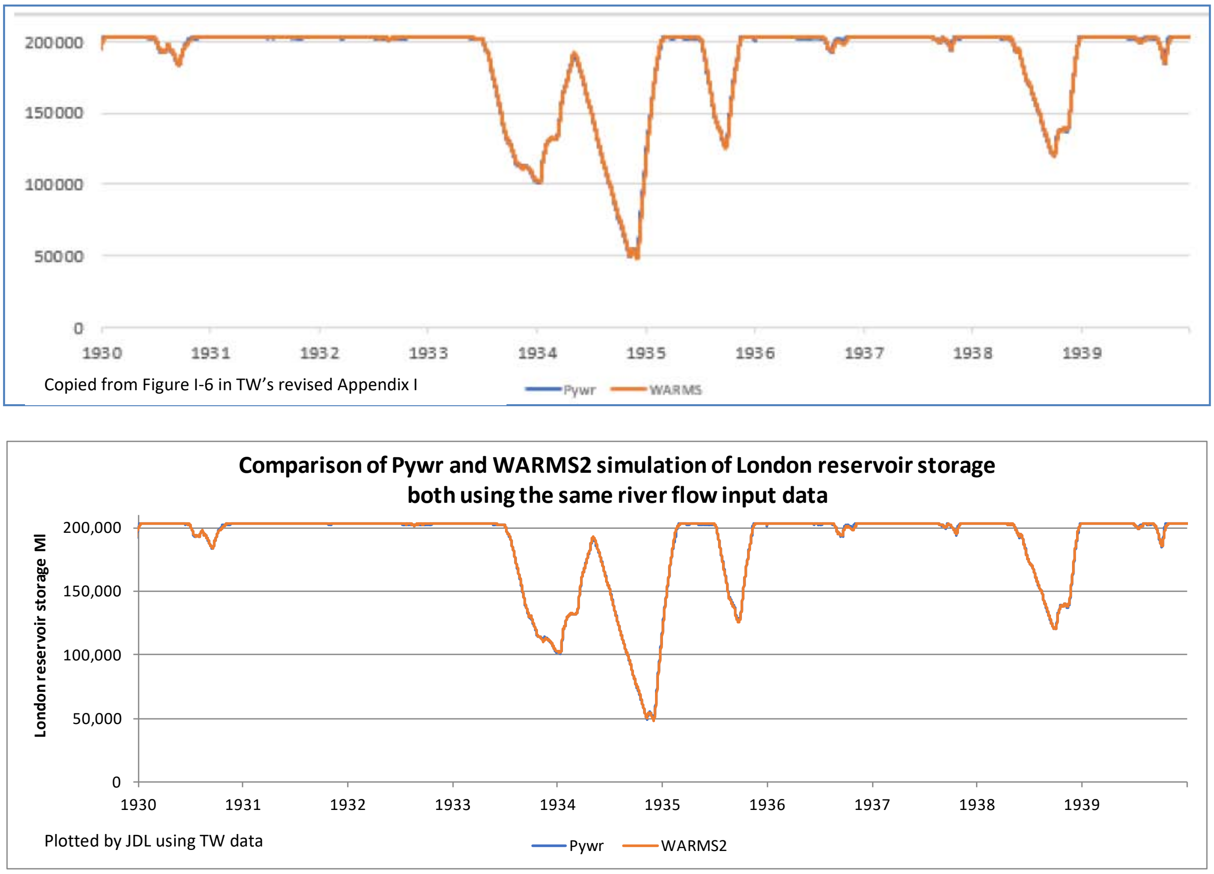


Figure 2 - Revised Step 1 validation of Pywr vs WARMS2 modelling using the same river flow data

On this basis, I agree that the up-dated Pywr simulation model provides an acceptable match to WARMS 2 simulation model output <u>when using the same river flow input data</u>. However, it is not clear whether TW included an out-of-date version of their Figure I-6 in their report or whether the out-of-date version of the Pywr model was used for all their work on the original draft WRMP.

2.2.2 Step 2 – Validity of the hydrological model used to generate river flows

The Step 2 model validation showed large differences in model outputs when historic river flows are generated from the historic weather data using the new hydrological model that was subsequently used to generate 19,200 years of daily flow records from the stochastically generated weather data.

Detailed evidence of this major flaw in Thames Water's modelling was presented on pages 12 to 17 of GARD's Addendum and is replicated in full below as Extract 1:

Extract 1 from GARD's Addendum to the dWRMP24 response

Step 2 of Thames Water's Pywr validation used river flows from generated from historic weather using the hydrological model that generated the 19,200 years of stochastic river flow data. The Pywr modelling is then compared with the WARMS model output which used river flows generated using different rainfall data from the same historic period and the WARMS hydrological model. The resulting plots in Figure I-7 of Appendix I show large differences in modelled storages in droughts as below:

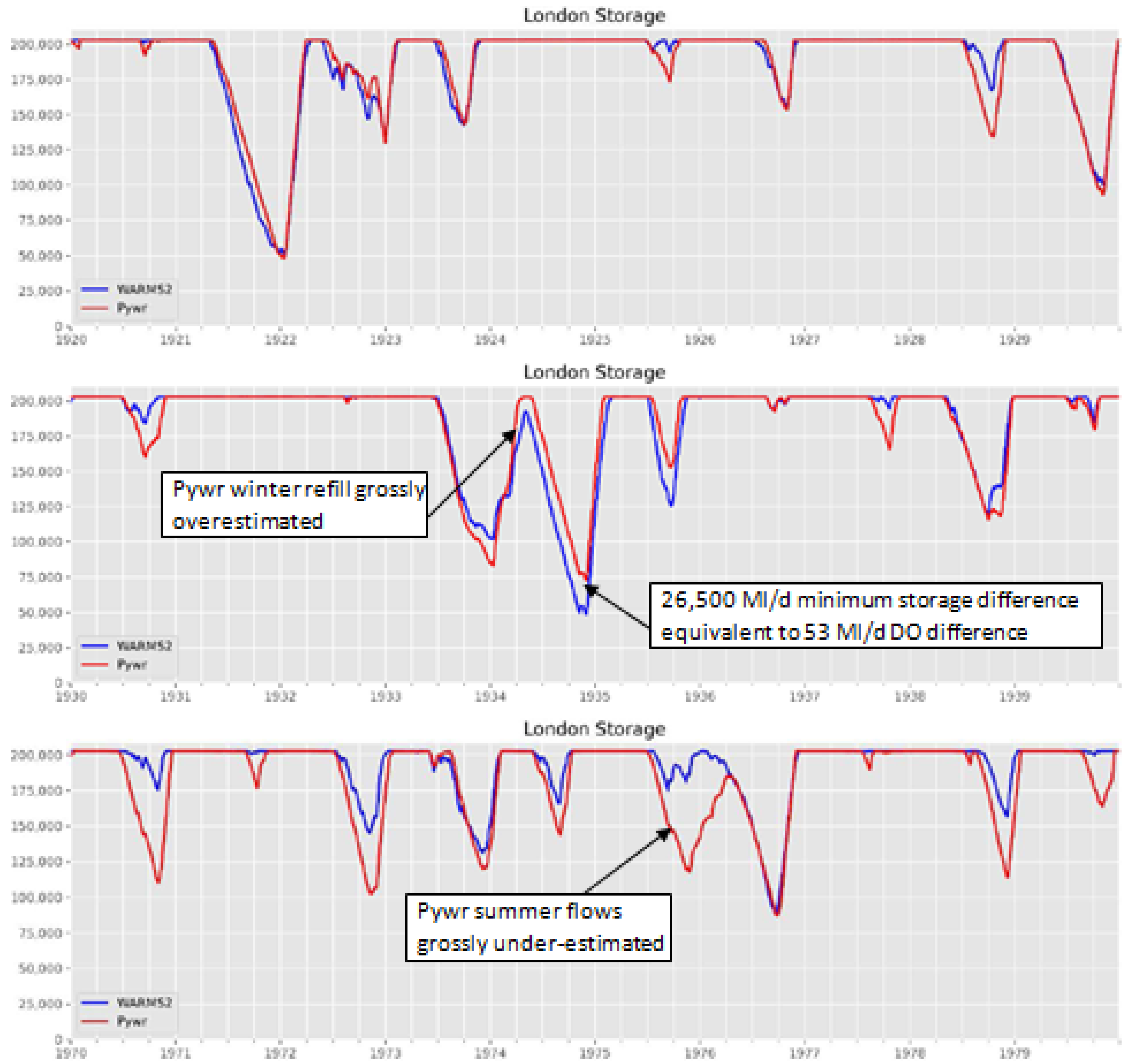


Figure I - 7: Step 2 Pywr Model Validation Plots

Figure 2 - TW Step 2 validation of Pywr, simulation of London storage using Pwyr flows

In Appendix I, Thames Water describes the validation fits above as follows[5]:

> *"Figure I - 7 shows validation plots for key drought periods for the fully updated*

[5] TW WRMP24 Appendix I, paragraph I.126

Extract 1 from GARD's Addendum continued …

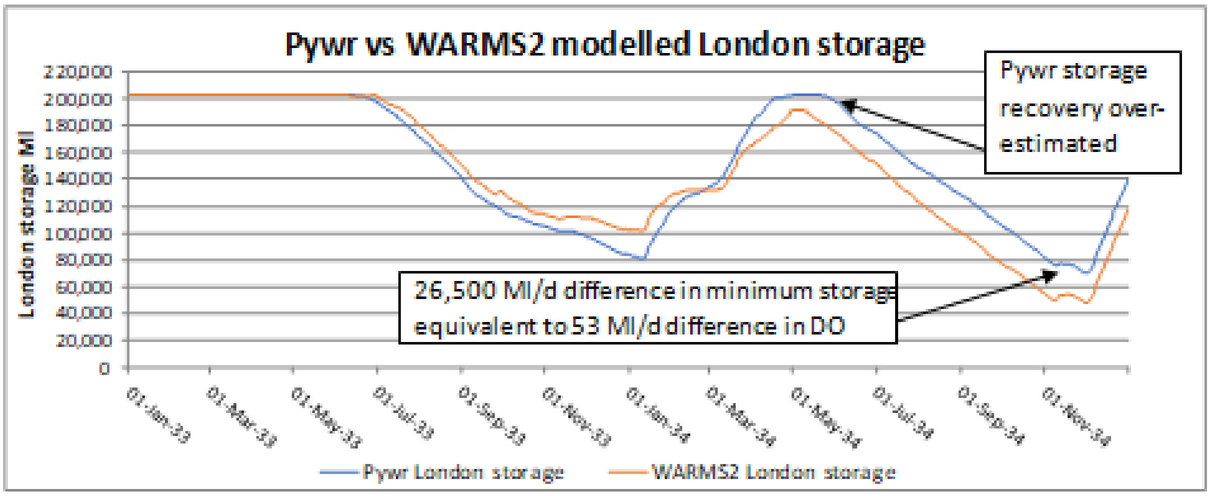
hydrological and water resources model (run in the 'WRSE North' configuration). These plots show a close agreement between Pywr and WARMS2 outputs for key drought periods, with the revised hydrological modelling/rainfall datasets seeming to suggest greater drawdowns during some moderately dry periods. The DO calculated when the model was run was 2296 Ml/d (a figure comparable with the 2302 Ml/d WARMS2 DO). Considering the degree of change that had been undertaken and results from WRMP19 hydrological modelling, this was considered a good fit."

The fits in the maximum drought drawdowns, which Thames Water describe as "close", are mostly poor. The Pywr modelled drawdown in the severe single-year drought of 1921 was similar to the WARMS modelled drawdown, enabling Thames Water to say that there was little difference in the deployable outputs shown by each model.

However, the Pywr maximum drawdown in the severe two-year drought of 1933-34 was 26,600 Ml less than the WARMS2 modelled drawdown[6]. This is equivalent to over-estimating the London deployable output by about 53 Ml/d (in two-year droughts, the London reservoirs take about 500 days to fall to minimum storage, so the deployable output difference is roughly the storage difference in Ml divided by 500 days).

The cause of this major difference between the Pywr and WARMS modelling of droughts is the differences in modelled Teddington 'natural' flows, as shown below for the historic 1933-34 drought[7]:

Extract 1 from GARD's Addendum continued ...



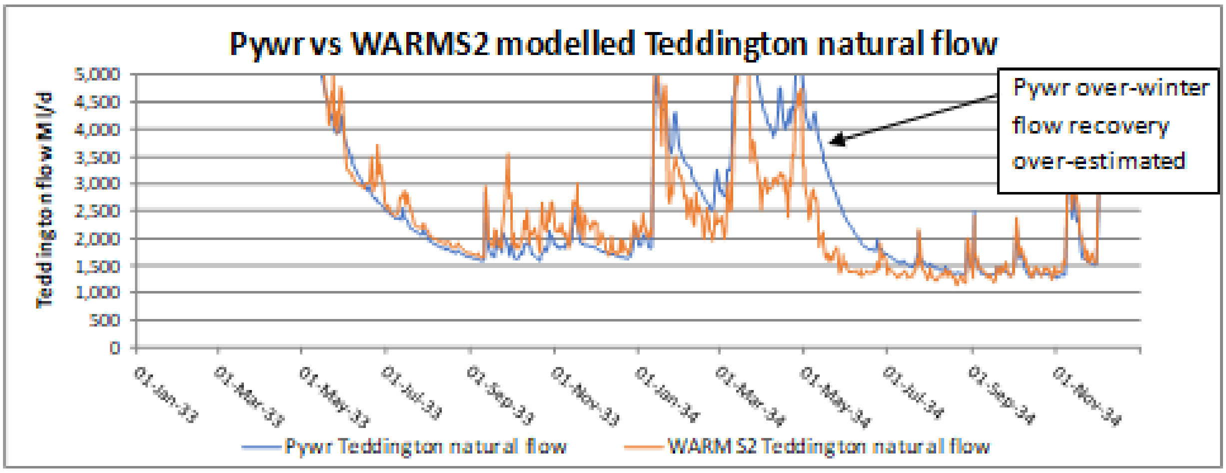


Figure 3 - Comparison of Pywr and WARMS2 modelling in 1933-34 drought

In the flows of the 2-year drought of 1933-34, Pywr over-estimates the over-winter river flow and London storage recovery, leading to a large over-estimation of residual storage and deployable output.

There is a similar picture in the major 2-year drought of 1943-44, which Thames Water did not show in Figure I-7 in Appendix I:

Extract 1 from GARD's Addendum continued …

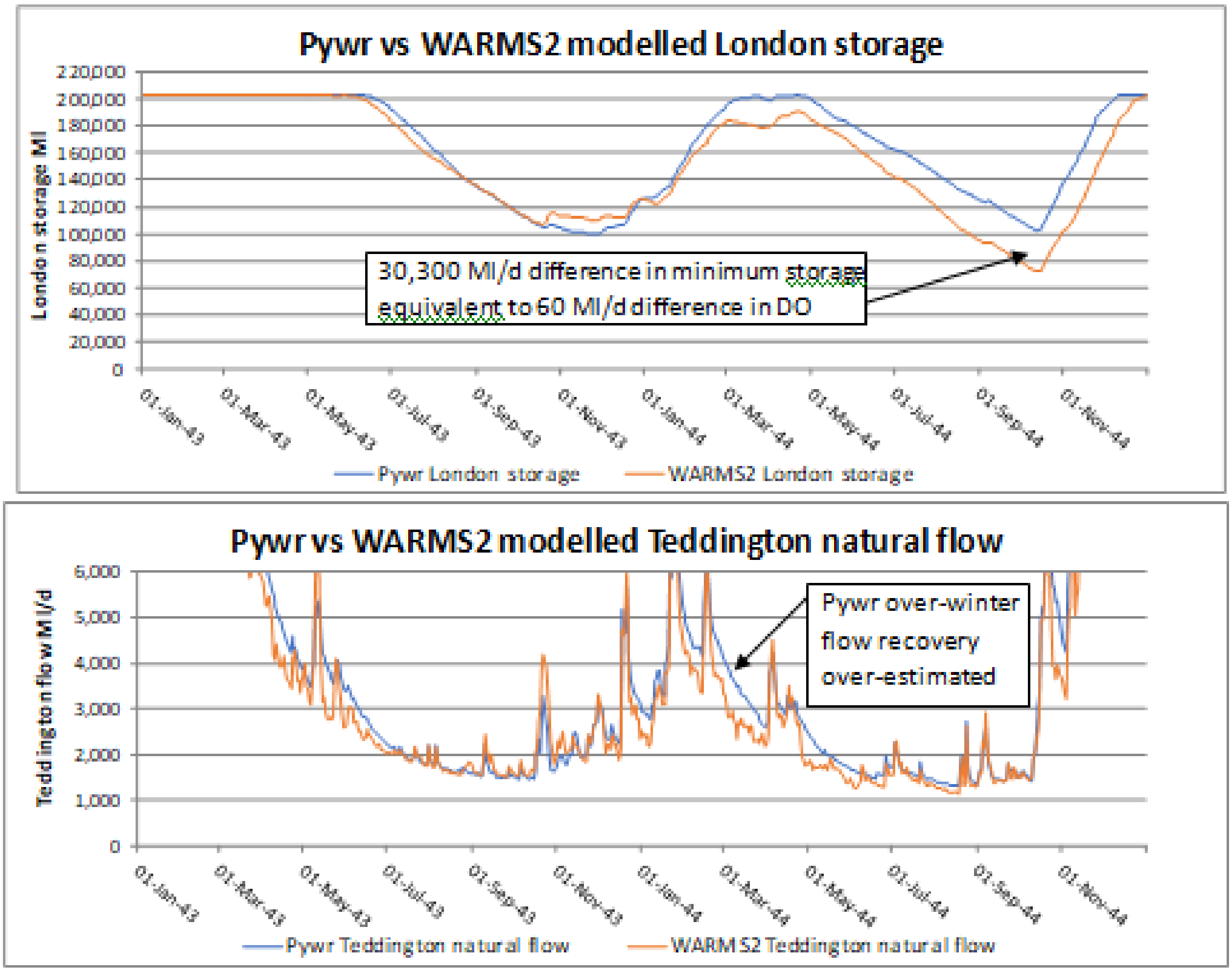
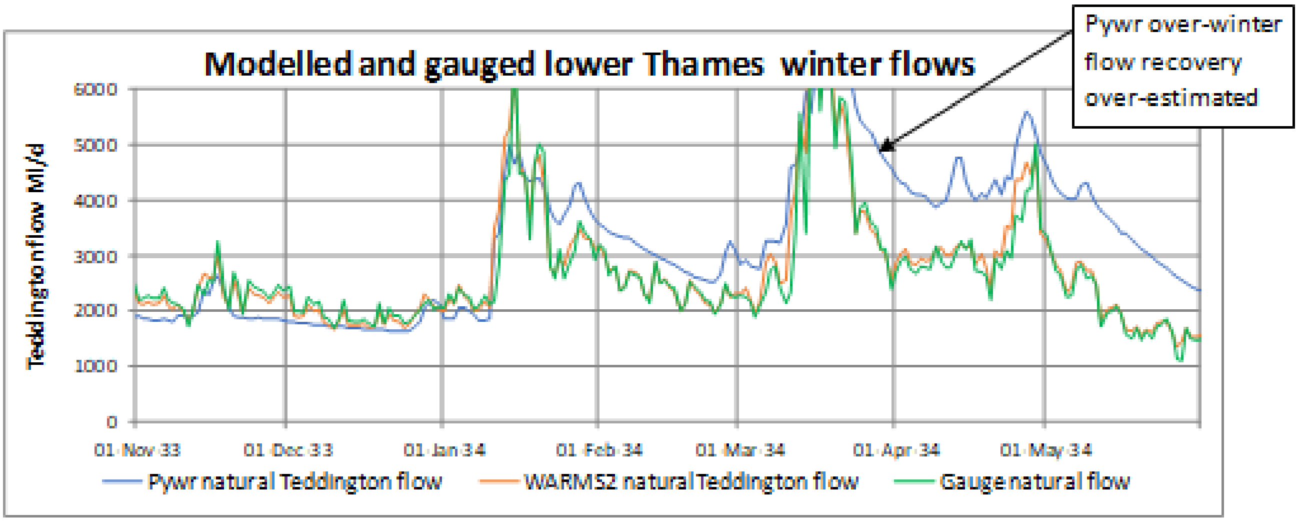


Figure 4 - Comparison of Pywr and WARMS2 modelling in 1943-44 drought

In the flows of 2-year drought of 1943-44, the Pywr model over-estimates the minimum storage by 30,300 Ml, equivalent to a deployable output difference of about 60 Ml/d. As for the modelling of the 1933-34 drought, the reason for the modelled storage differences is the differences in modelled natural flows during the winter between the two summer droughts.

The differences in modelled winter flows during 2-year droughts can be seen more clearly in Figure 5, which also shows the gauged naturalised Kingston flows, which the NRFA web-site explains are derived from the actual gauged flows by adding back all the lower Thames abstractions by Thames Water and Affinity Water[8]:

Extract 1 from GARD's Addendum continued …



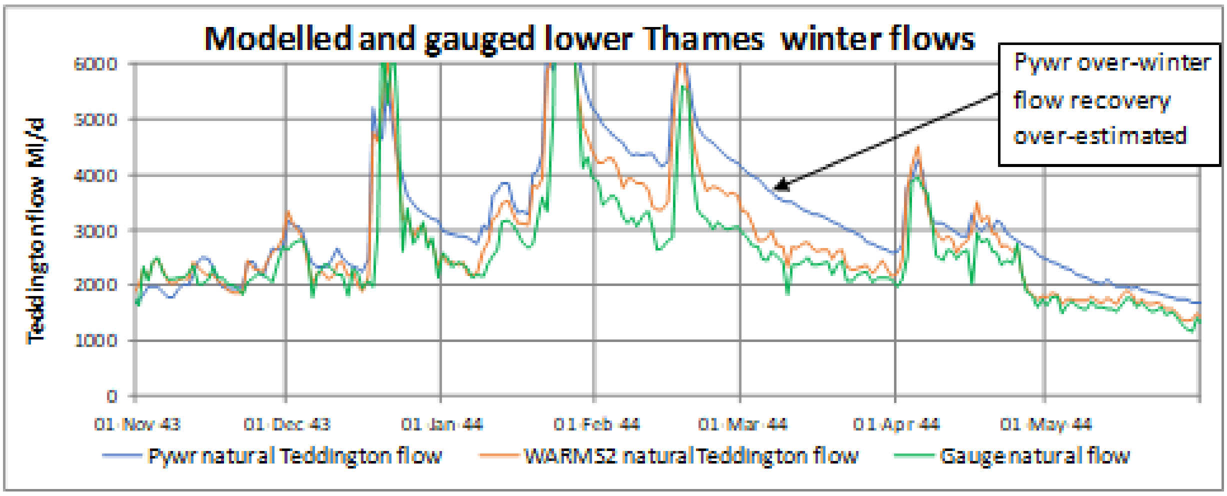


Figure 5 - Modelled and gauged natural winter flows in 2-year droughts

The Pywr model grossly overestimates the winter flow recovery during the 2-year droughts of 1933-34 and 1943-44. The WARMS2 modelling of the naturalised flows is a much better fit to the naturalised gauged flows, although there is some over-estimation of flow recovery in the winter of the 1943-44 drought.

Although the Pywr used a different version of historic rainfall data to that used in the WARMS model[9], this analysis suggests that the hydrological model used to generate river flows from climatic records provides a poor simulation of the Thames catchment response to winter rainfall occurring at the end of a long summer drought. It appears that the modelled flows used in the Pywr model may respond too quickly to the winter rainfall, not taking sufficient account of the need for groundwater level recovery before flows can recover in the chalk tributary catchments which dominate the pattern of flows in the Thames. This would have profound implications for assessing the deployable output of the Abingdon reservoir and STT options:

Extract 1 from GARD's Addendum continued …

- For Abingdon reservoir, the over-estimation of winter flow recovery would disguise the reservoir's lack of resilience in long duration droughts.

- For the Severn to Thames transfer, over estimation of winter flow recovery in the Thames would diminish the benefit of the unsupported transfer, when flows in the Severn recover faster than flows in the Thames, as is always the case when droughts end.

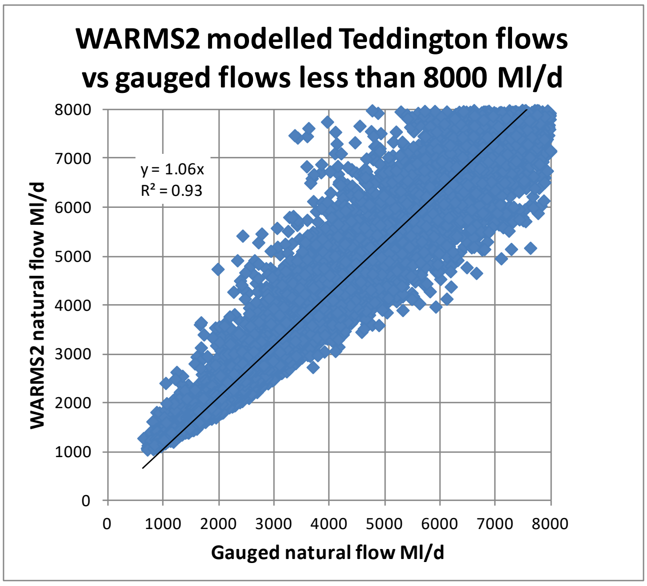
The implications of this on the deployable output and drought resilience of the Abingdon reservoir and STT options are further discussed in Sections 3 and 4 of this Addendum.

Thames Water response to the detailed criticism that we presented in Extract 1 from the Addendum starts with the extract shown below:

From page 106 of Appendix G2 of TW's SoR

We do not agree that the hydrological models used in our stochastic modelling "grossly overestimate" winter flow recovery in their validation and as such we do not agree with the conclusions drawn in respect of impacts of over-estimation of SESRO DO values and under-estimation of STT DO values. The hydrological models used have Nash-Sutcliffe Efficiency (NSE) and logNSE flow statistic values (commonly used flow statistics to ascertain hydrological model performance, with logNSE particularly applicable for low flow prediction) of around 0.9, indicating strong performance.

As a test of the validity of TW's claim, we have shown below scatter plots of Pywr modelled flows and WARMS2 modelled flows against the naturalised gauged flows below 8000 Ml/d (the significant flows for filling London's reservoirs) for the historic period 1920 to 1997:



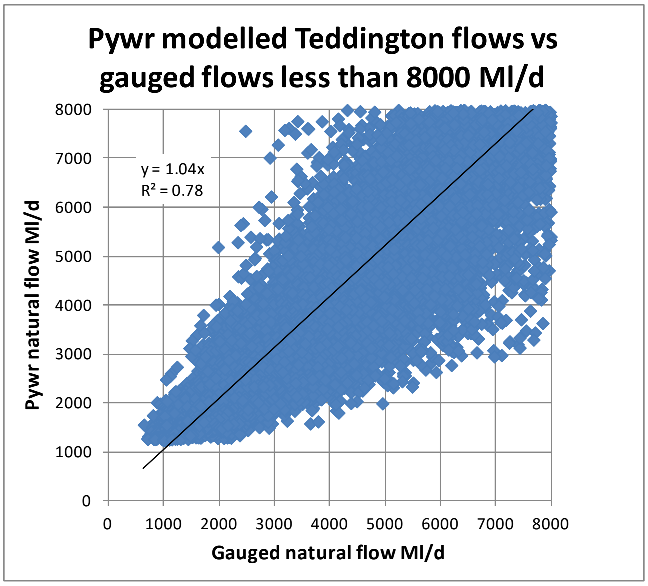


Figure 3 - Scatter plots showing fits of Pywr and WARMS2 flows against gauged flows

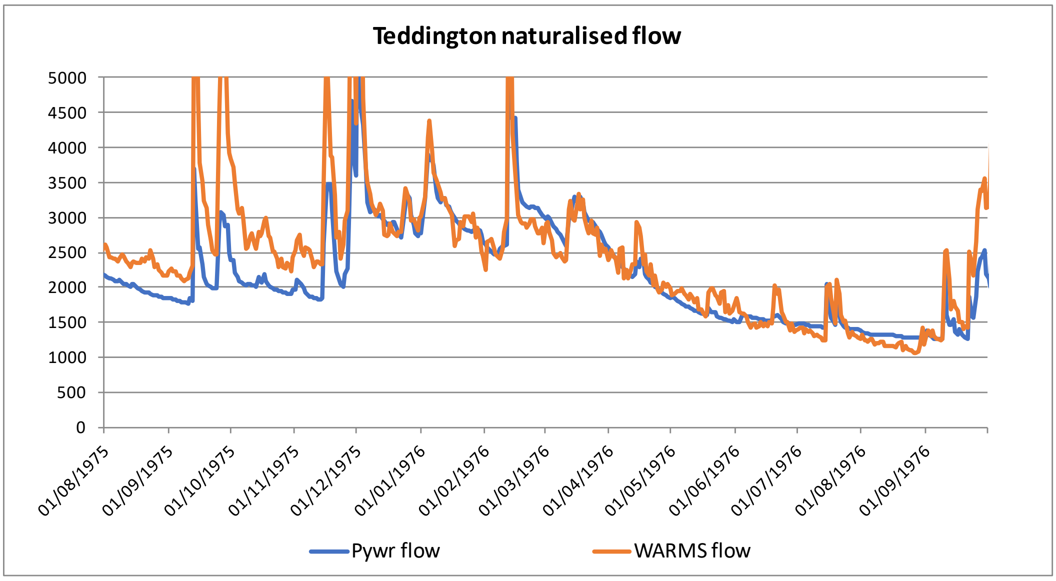
The plots in Figure 3 show that the Pywr natural Teddington flows are a poor fit to the gauged flows and a much worse fit than the WARMS2 flows. The R^2 value of 0.78 for the Pywr plot is much lower than the Nash-Sutcliffe Efficiency value of 0.9 quoted by Thames Water (R^2 values are usually similar to NSE values). It is hard to see how Thames Water can justify the fit between the Pywr flows and gauged flows as "indicating strong performance".

Thames Water then goes on to dismiss GARD's arguments about the inadequacy of the Pywr flows as follows:

From pages 106-107 of Appendix G2 of TW's SoR:

While we agree that WARMS2 flows are a closer match to gauged naturalised records, that is because WARMS2 uses gauged naturalised flows as an input. In WARMS2, the gauged naturalised flows are used to estimate flow contributions from ungauged catchments and those with "fast" hydrological responses. It would not be possible to use only the hydrological models in WARMS2 on their own for use in stochastic modelling.

GARD have presented evidence from the events of 1933-34 and 1943-44 to demonstrate events where flow over-prediction during winter occurs. The figure below demonstrates that during 1975-76 (an event which GARD make much of in their later analysis), the hydrological models used in our flow modelling under-predict winter flows, compared to WARMS2. This demonstrates that it is not the case that the hydrological models used in our stochastic modelling systematically over-estimate flows during winter recharge periods, rather it is that hydrological modelling is challenging, perfect model validation cannot be achieved, and thus that variance between observed and models flows, and between flows produced by different models, are to be expected. Our consideration is that our hydrological models are fit for purpose.



Note graph has been re-plotted to get a clearer image, but is identical to TW's plot

Although I agree that the plot above shows that the Pywr model does not always overestimate winter flow recovery, nevertheless in the two most severe 2-year droughts in the past 100-years, the Pywr modelling does substantially over-estimate winter flow recovery. It therefore seems inevitable that the Pywr modelling will substantially over-estimate flow recovery in the intervening winters in many of the severe 2-year droughts in the 19,200 stochastic record.

Furthermore, TW's plot above of modelled flows in the winter of 1975-76 doesn't show the full picture of the inadequacy of Pywr modelling of the drought of 1975-76, which I have generated below:

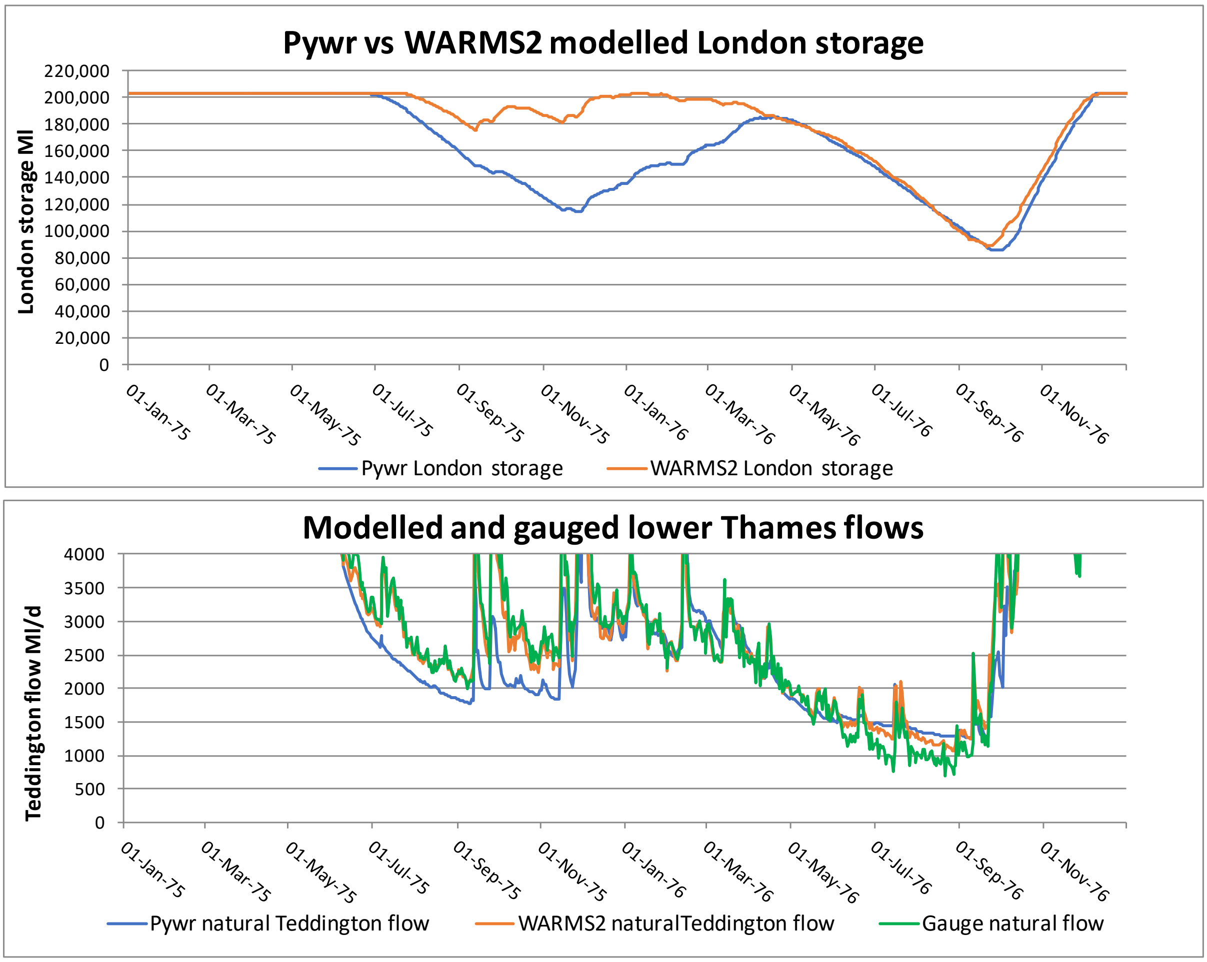


Figure 4 – Pywr and WARMS2 modelling of the historic 1976 drought

The Pywr modelling of reservoir storage in the first half of the 1975-76 drought shows little resemblance to the WARMS2 modelling, noting that the lower plot shows that the WARMS2 Teddington flows are a much better fit to the gauged flows in 1975-76. With the large mis-match in the modelling of the first year of the drought, it is only by chance that the Pywr modelling almost matches the WARMS2 minimum storage in September 1976. This is further evidence of the unreliability of the Pywr simulations of the long duration droughts which are crucial to determination of the deployable output and drought resilience of Abingdon reservoir.

2.2 Validity of the 19,200 years of stochastic river flows

2.2.1 GARD's concerns over stochastically generated natural Teddington flows

GARD has major concerns about the stochastically generated 19,200 years of Teddington natural flows used in the Pywr modelling. I have reproduced these concerns in full in Extract 2 below, copied from GARD's Addendum.

Extract 2 from GARD's Addendum pages 17-21

2.2 Validity of stochastically generated river flows

Overall concerns

In Section 2.1, we have demonstrated the flaws in the hydrological modelling used to generate river flows from historic climate records. In particular, the generated river flows greatly over-estimate the speed and amount of flow recovery in the intervening winters during the 2-year droughts in which London's supplies are most vulnerable. These flaws have led to the poor validation of Pywr modelling when compared to WARMS2 modelling using historic flows during the 2-year droughts of 1933-34 and 1943-44. This is shown by Thames Water's own analysis in Figures 1 and 2 and our analyses in Figures 3 to 5.

In addition, we continue to have major concerns about the use of the 48 year period 1950 to 1997 as the basis for generating 19,200 years of stochastic river flows, excluding the severe droughts of 1921, 1933-34 and 1943-44. We expressed these concerns in our response in October 2020 to WRSE's consultation on their Method Statements for preparing their regional plan, but we can see no evidence in Thames Water's WRMP that our concerns have been considered or acted upon.

In short, the use of historic climate data only for 1950-1997 means the exclusion of:

- the droughts of 1921, 1933-34 and 1943-44, the three most severe droughts of the past 100 years for London's supplies
- the past 25 years of most rapid climate change

Therefore, the period 1950-1997 is an unsatisfactory basis for generating the stochastic data, lacking both the period of extreme low flows pre-1950, with several long duration droughts, and the recent period of most rapid climate change.

Exclusion of droughts of 1921, 1934 and 1944

The exclusion of the droughts of 1921 and 1933-34 is of particular concern, because they both extended well into the winter and London reservoir storage would not have started to recover until 20th January 1922 or 5th December 1934 respectively. The most severe drought

Extract 2 from GARD's Addendum continued …

in Thames Water's selected history, 1950-1997, was 1975-76 in which reservoir levels started to recover on 23rd September 1976. The period 1950-1997 contained no severe droughts extending into late autumn/winter like 1921 and 1934, so this type of long drought is not adequately represented in the 19,200 year stochastic record. We also note that the drought of 2011-12 which extended to early 2012 is excluded from the historic base period.

The 19,200 years of stochastic river flows are generated by perturbing the weather patterns of the period 1950 to 1997. It is, therefore, inevitable that the droughts generated in the 19,200 year stochastic flow record have followed the pattern of droughts in the historic period 1950 to 1997. This is shown by the following plot taken from the report on the WRSE Regional System Simulator[10] and described *as "a heat map to more easily understand where failures occurred across the stochastic data set"*:

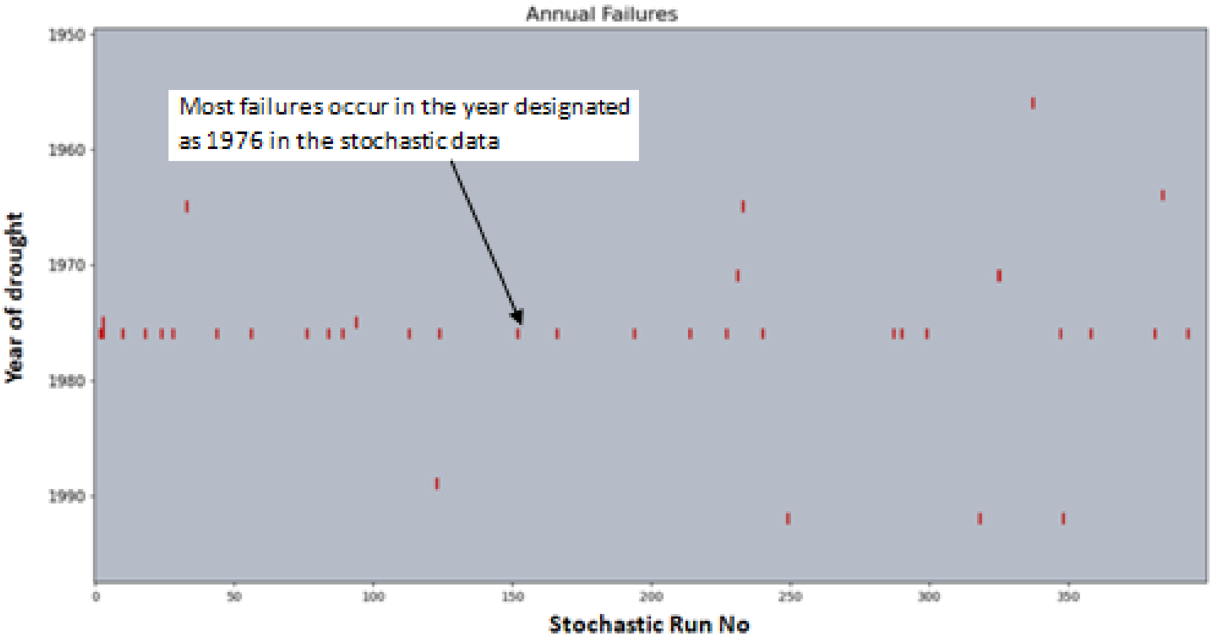
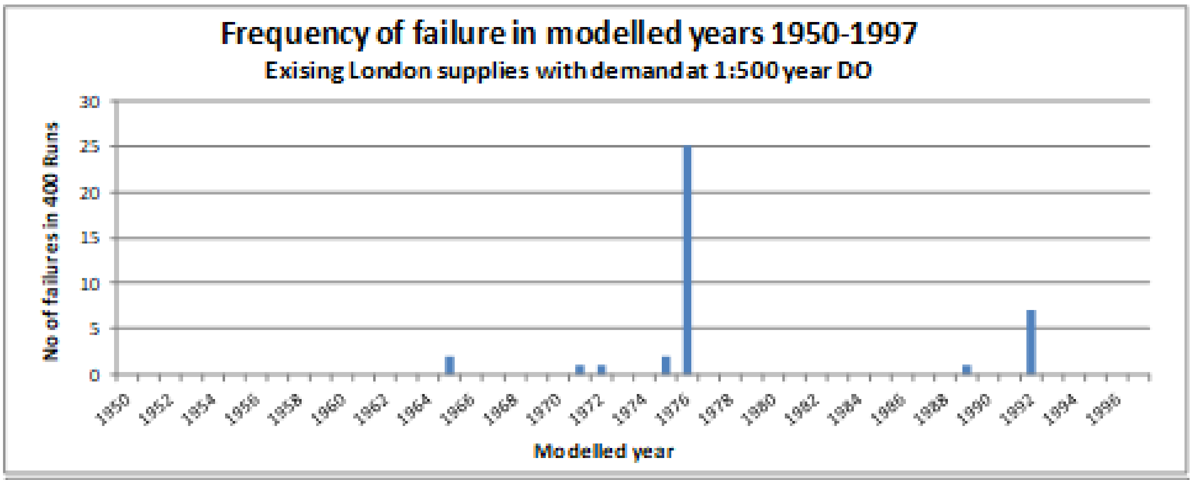


Figure 6 - 'Heat map' showing distribution of severe droughts in the 19,200 year record

This plot shows that 27 of the 37 droughts plotted on Figure 6 occurred in the modelled years 1975-76. The remaining 10 droughts occurred in lesser droughts in the historic record, like 1992. The Regional System Simulator report does not explain how the above plot was derived, but we have generated similar patterns by analysing the 19,200 years of Level 4 (ie supply failure) control line crossing data provided under EIR 22-23-390 for existing London supplies[11] and London supplies with the 150 Mm^3 Abingdon reservoir[12], as shown in Figure 7:

[10] WRSE Regional System Simulator, August 2021, Atkins, Figure 3-3
[11] Data from EIR-22-23-390 file *'tw-london-stochastic-baseline-v5 last day dy-failures.csv'*
[12] Data from EIR-22-23-390 file *'tw-sesro-150-london-stochastic-baseline-v8 last day dy-failures.csv'*

Extract 2 from GARD's Addendum continued …



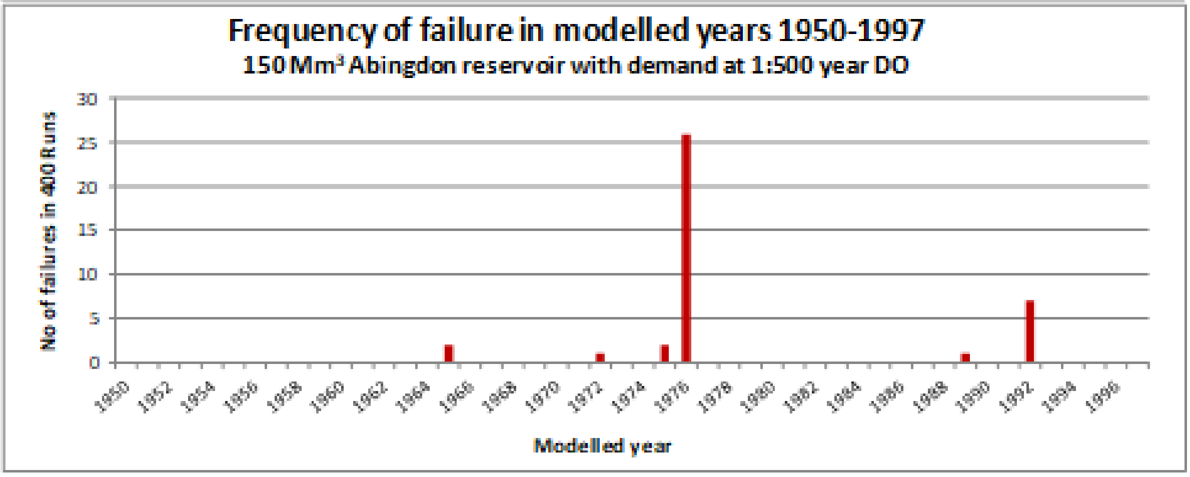


Figure 7 - Pywr modelled frequencies of supply failure in each year of stochastic data

The patterns of 'year of failure' are almost identical in the Pywr 19,200 year simulations of existing London supplies and existing supplies with Abingdon reservoir, in each case replicating the pattern of historic droughts in the period 1950-97 – much the most severe drought was 1975-76. There are also some modelled supply failures in stochastic flows for 1992, which was not a significant historic drought year, but followed three consecutive dry years in 1989 to 1991 – see later comments following Figure 9.

It is evident that the method of generating stochastic river flows has retained the general pattern of historic flows 1950 to 1997, varying the intensity of droughts whilst keeping their general shape and duration. The method will not generate droughts of different shapes, for example droughts of much longer duration. This danger was identified in WRSE's Method Statement for Stochastic Climate Datasets[13]:

> *"As with any dataset generated based on existing datasets using statistical methods, the stochastic weather sequences are only as good as the datasets on which they are trained. As stated above, the stochastic dataset is formed of 400 48-year sequences and*

[13] WRSE Method Statement on Stochastic Climate Datasets: Consultation Version, July 2020, paragraph 2.7

Extract 2 from GARD's Addendum continued …

is trained on the 1950-1997 baseline period. There is a risk that extreme, extended droughts may not necessarily be well reflected in the dataset, although quantifying this risk is extremely difficult. Companies may complement the stochastic dataset with drought artificial weather series to represent prolonged drought events (which the stochastic generator will not have been trained on)."

The relative severity of actual droughts of the past 100 years, in terms of their impact on London's supplies, can be seen from the plot below of minimum modelled drawdown, derived from Thames Water's modelling of existing London supplies at a deployable output of 2302 Ml/d[14]:

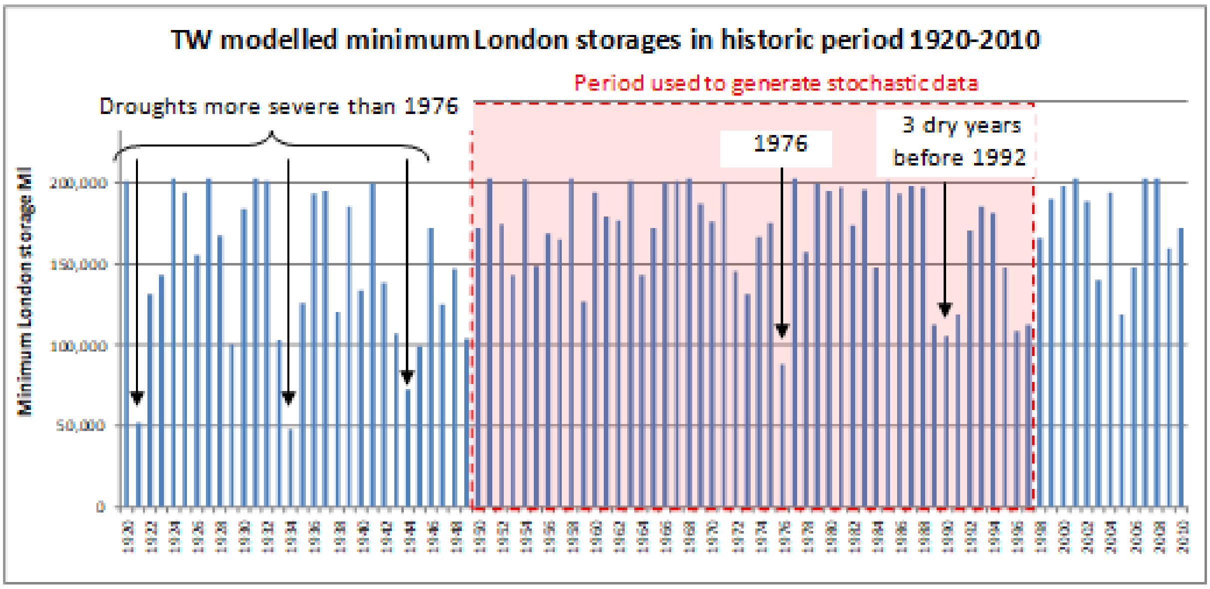


Figure 8 - Relative severity of droughts 1920-2010 as impact on minimum London storage

As can be seen in Figure 8, the droughts of 1921, 1934 and 1944 were all a lot more severe than the 1976 drought in terms of impact on London reservoir storage. The reason for the greater impact of the 1921, 1934 and 1944 droughts was the <u>length</u> of the droughts, with the longest droughts causing the greatest drawdown of London storage, as shown in Figure 9[15]:

[14] Data from EIR-22-23-390 file *"AR20 Q4 Sc1 - London DO – 2302.xlsx"*
[15] Ibid

Extract 2 from GARD's Addendum continued …

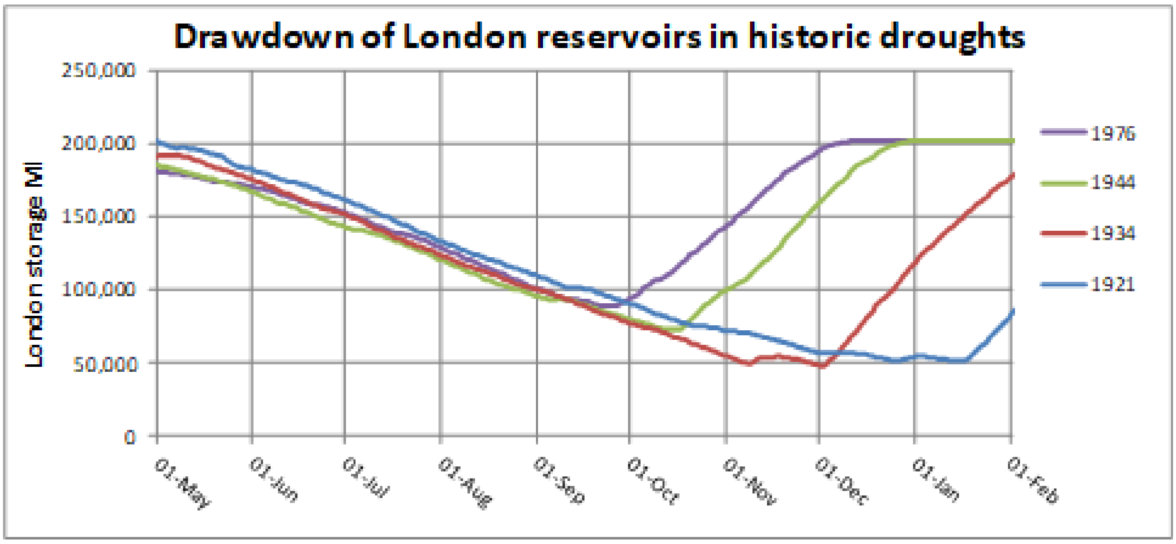


Figure 9 - TW modelled drawdown of London reservoirs in major historic droughts

It is evident from this plot that the 1976 drought was of much too short a duration to be a representative 'training drought' for all the stochastically generated droughts. The absence of any other significant droughts in the 1950 to 1997 period further limits the variability of the stochastically generated droughts.

The potential significance of much longer droughts, or sequences of dry years, is shown by number of supply failures generated by the stochastic modelling in 1992, as shown earlier in Figure 6. Historically, 1992 was not a particularly dry year, as can be seen from the modelled historic drawdowns in Figure 8. However, the presence of three moderately dry years in 1989 to 1991 has evidently been enough for the hydrological modelling to generate some very severe droughts in 1992.

As we will show later, the 1992 droughts in the stochastic record are much longer droughts than those occurring in 1976 and they test the resilience of Abingdon reservoir, typically emptying the reservoir well before the end of the drought. However, because the period 1989 to 1992 was not exceptionally dry, there are relatively few droughts of this type in the stochastic record. As we shall show in Section 2, there were more droughts of this type in the stochastic data for WRMP19, which were 'trained' on the period 1920 to 1997, ie including the 1921, 1933-34 and 1943-44 droughts. We suspect that there would have been still more of this type of long drought if weather data for the 1890s had been included in the 'training period'.

Despite this problem being identified in WRSE's 2020 Method Statement, it appears to have been ignored in Thames Water's WRMP. We have found no evidence that Thames Water have followed WRSE's suggestion of "*Companies may complement the stochastic dataset with drought artificial weather series to represent prolonged drought events (which the stochastic generator will not have been trained on)*". For example, what would happen if a drought like 1976 was to follow a drought like 1921, 1934 or 1944?

2.2.2 TW's response to GARD's criticism of stochastic Teddington flows

Thames Water's response to this detailed criticism of stochastic Teddington flows failed to address most of the evidence presented in Extract 2 above and focused only on the validity of the stochastic weather data. They have not properly addressed GARD's key point, which is that the stochastic data generated different versions of the patterns of rainfall and drought sequences in the historic period 1950 to 1997, so was unable to generate the types of long drought, of more than 18 months duration, against which Abingdon reservoir has poor resilience.

All that TW's response says about possible severe droughts longer than 18 months is this:

From page 115 of Appendix G2 of TW's SoR:

If there is a 1 in 500-year drought event of 18 months duration, while longer droughts could occur, they would either be:

- *Equally dry (in terms of mm rainfall per month) for a longer duration, and thus more severe than a 1 in 500-year event (each month of very low rainfall being an unlikely event)*
- *less dry (in terms of rainfall per month) and thus not as much of a risk for the existing London supplies*

Thames Water provides no evidence to support the statement above and it appears to be merely a subjective view formulated to get round the inconvenient fact that modelling shows that Abingdon reservoir would be virtually empty at the end of most major 18-month droughts and so would be unable to provide any more water in longer droughts.

GARD's concern about longer droughts is shared by WRSE in their method statement on the stochastic climate datasets[1]:

> *"There is a risk that extreme, extended droughts may not necessarily be well reflected in the dataset, although quantifying this risk is extremely difficult. Companies may complement the stochastic dataset with drought artificial weather series to represent prolonged drought events (which the stochastic generator will not have been trained on)."*

Thames Water has rejected the suggestion of using artificial weather series to represent prolonged drought events by saying this:

From page 115 of Appendix G2 of TW's SoR:

As also noted in a previous response, what GARD refer to as "advice" from WRSE is not advice, and is instead an allowance from WRSE to diverge from the preferred methodology should companies consider that a key vulnerability of their supply system is omitted. Our consideration is that the stochastic datasets properly consider drought events which may occur and to which our supply system is vulnerable.

[1] WRSE Method Statement on Stochastic Climate Datasets: Consultation Version, July 2020, paragraph 2.7

Again, this is merely Thames Water's opinion, with no supporting evidence, presented as an excuse for not looking at long droughts that Abingdon reservoir cannot cope with. I have addressed this further in Section 3.6 of this paper.

Thames Water's comment's on the validity of the stochastic weather data used to generate Teddington natural flows started with this:

From page 107 of Appendix G2 of TW's SoR:

We do not agree that the stochastic weather datasets adopted in our WRMP24 modelling of Deployable Output are not fit for purpose. The datasets are shown in the WRSE technical note accompanying their production (Regional Climate Data Tools Final Report) to provide a good fit to historical data (when considering different rainfall deficit durations, and when considering the historical period 1920-1997, as shown in Section B.4.3 of the WRSE/Atkins report), the stochastic datasets have been widely adopted across the UK Water Industry, and the Deployable Output figure for 1 in 100-year Deployable Output is close to the figure from our existing "Worst historical" (1920-2013) modelling.

My response to these comments is:

1. The fact that the stochastic data sets have been widely adopted across the UK water industry, doesn't make them any more valid.

2. I agree that there is some reassurance from the 1:100 year deployable output of existing supplies using the stochastic data matching the modelled DO in the worst historic drought. However, that doesn't address GARD's main point which is that the method of generating the stochastic flows greatly reduces the likelihood of the type of multi-year drought (eg 3 dry summers with two intervening dry winters) in which Abingdon reservoir has poor resilience.

3. Although the rainfalls generated by the 1950s model are similar to the data generated from the 20th century model, both sets of stochastic data are often widely different to the observed data, particularly at the 1:500 year return period.

4. The "good fit to historic data" which Thames Water says is demonstrated in Section B.4.3 of the WRSE/Atkins report is a very subjective view, which, in my opinion is not supported by the data published in that report. Section B.4.3 includes droughts from several regions, so I have extracted below some data relating to the South-East[2], highlighting in yellow where I consider the fits to be poor:

[2] Copied from Section B.4.3 of WRSE/Atkins Final Report on Regional Climate Data Tools, July 2020 https://www.wrse.org.uk/media/ok1mtsoq/wrse_file_1338_regional-climate-data-tools.pdf

Metric	Region	RP	Obs	20th Century model	1950s model	Diff 20thC – 1950s (per month)	Diff Obs – 20thC (per month)	Diff Obs – 1950s (per month)
April - August	WRSE	50	116.3	158.4	147.4	2.2	-8.4	-6.2
	WRSE	100	81.6	144.5	132.4	2.4	-12.6	-10.2
	WRSE	200	47.2	132.3	119.4	2.6	-17.0	-14.4
	WRSE	500	1.6	118.2	104.3	2.8	-23.3	-20.5
October - September	WRSE	50	501.7	551.8	n/a	n/a	-4.2	n/a
	WRSE	100	459.8	516.6	n/a	n/a	-4.7	n/a
	WRSE	200	423.6	485.5	n/a	n/a	-5.2	n/a
	WRSE	500	381.6	448.9	n/a	n/a	-5.6	n/a
January - December	WRSE	50	565.7	559.1	n/a	n/a	0.6	n/a
	WRSE	100	514.3	528.9	n/a	n/a	-1.2	n/a
	WRSE	200	463.4	502.7	n/a	n/a	-3.3	n/a
	WRSE	500	396.3	472.3	n/a	n/a	-6.3	n/a
October - March	WRSE	50	260.1	273.9	280.8	-1.1	-2.3	-3.4
	WRSE	100	237.7	254.5	262.3	-1.3	-2.8	-4.1
	WRSE	200	218.2	237.6	246.3	-1.5	-3.2	-4.7
	WRSE	500	195.5	218.1	227.7	-1.6	-3.8	-5.4
November - February	WRSE	50	152.4	170.6	167.1	0.9	-4.5	-3.7
	WRSE	100	132.7	155.2	151.8	0.9	-5.6	-4.8
	WRSE	200	115.1	141.8	138.3	0.9	-6.7	-5.8
	WRSE	500	94.1	126.5	122.5	1.0	-8.1	-7.1
18 months to September	WRSE	50	823.5	865.2	859.1	0.3	-2.3	-2.0
	WRSE	100	774.1	828.2	820.5	0.4	-3.0	-2.6
	WRSE	200	728.8	796.2	787.0	0.5	-3.7	-3.2
	WRSE	500	673.2	759.1	748.0	0.6	-4.8	-4.2
24 months to September	WRSE	50	1270.3	1258.8	1263.6	-0.2	0.5	0.3
	WRSE	100	1205.9	1212.3	1217.5	-0.2	-0.3	-0.5
	WRSE	200	1142.2	1172.0	1177.4	-0.2	-1.2	-1.5
	WRSE	500	1058.2	1125.3	1130.9	-0.2	-2.8	-3.0

Note: i). Rainfall in mm
ii). Data fits considered poor indicated thus: [dashed red box]

Table 1 – WRSE/Atkins validation of stochastically generated rainfall

5. The data shown in Table 1 are only for rainfall – there are likely to be bigger differences in long return period river flows due to the flaws in the hydrological model that converts the stochastically generated rainfall into river flows, as described in Section 2.1.2 of this paper.

Thames Water then go on to describe how the stochastic data were derived and they repeat their claim that the stochastic rainfall records provide a good match to historic records:

From pages 107-108 of Appendix G2 of TW's SoR:

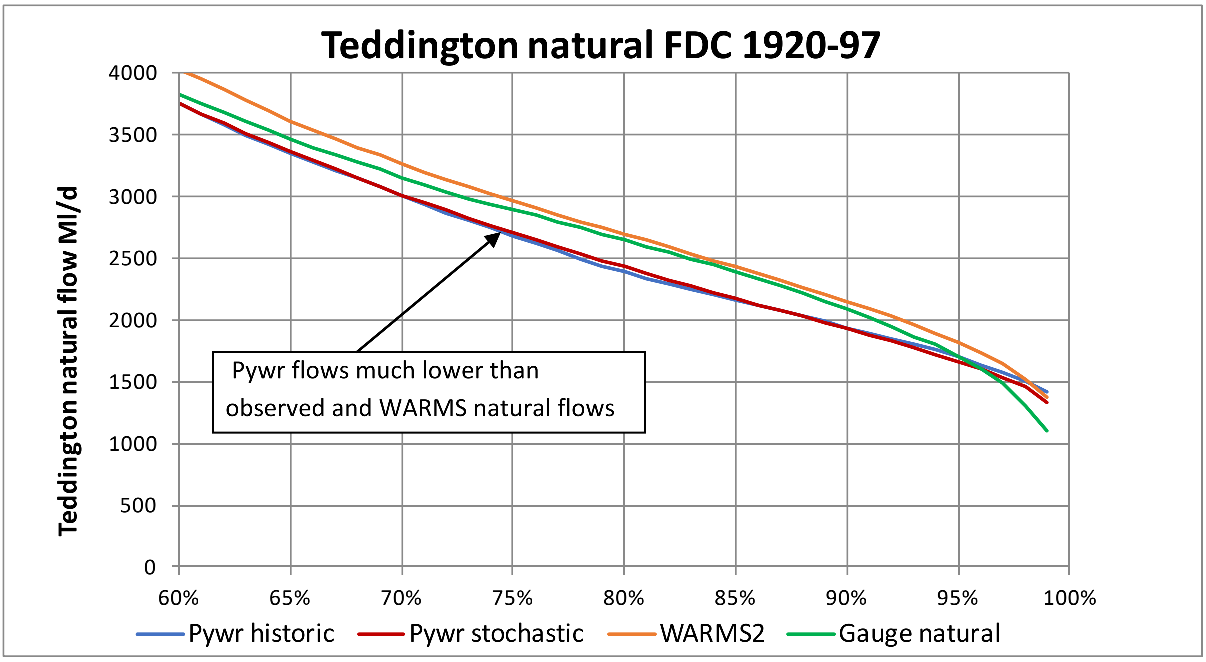
The stochastic weather datasets are generated using statistical processes in which rainfall is linked to climate drivers (e.g., the North Atlantic Oscillation Index, or sea surface temperature anomaly indices), but including an element of randomness. Statistical models are fitted based on "training" datasets of these climate drivers and rainfall volumes. The trained models are then used to generate different versions of what rainfall could reasonably have fallen when considering the climate drivers over that period, and considering the semi-random nature of rainfall. In the generation of stochastic rainfall datasets, historical rainfall datasets are not reproduced, and rather different rainfall time series are produced which represent what could have happened. As can be seen from the results in the WRSE/Atkins report, and as is explicitly stated in that report, the stochastic rainfall records provide a good match to the historical records, when considering the historical record of the full twentieth century.

Furthermore, recognising the potential criticisms which GARD now raise (i.e., that the first half of the twentieth century contained three severe droughts and thus a rainfall generator based on only the second half of the twentieth century would not adequately produce drought events), in the stochastic datasets project, a model was trained using a training period which also included the early part of the 20th century (noting that, for the reasons noted in the next paragraph, this model was compromised in terms of climate-driver data availability).

The report includes the statement that "The analysis concluded that while the [training set of the] 1950s [1950-1997] model does not include some of the key droughts in the 20th Century, in most cases this model performed as good [well] as, or marginally better [than the model trained on the longer 20th Century period], when viewed against the observed data in the 20th Century", i.e., the model used in the production of our WRMP24 stochastic rainfall datasets provides as good as, or a better match for the rainfall of the whole 20th Century than a model trained on data from the whole 20th Century.

Thames Water says that their consultants explicitly state a good match between historic and stochastic rainfall as though it shows an established fact, whereas it is actually their consultants' subjective judgment of their own work. I accept the point that the cumulative stochastic rainfall of different return periods is similar for data trained on 1950-1997 and data trained for 1920 to 1997. However, both sets of rain data are at times substantially different to the observed historic data, as can be seen in the WRSE/Atkins table that is reproduced in my Figure 1 above.

The differences between the stochastic modelled data and the observed historic data can also be seen in the Teddington river flow duration curves below, focusing on the natural flow below 400 Ml/d during which the London reservoirs are rapidly drawn down:



The Pywr stochastic data are from the full 19,200 year record in file "Teddington weir stochastic flow GR6J"

Figure 5 – Modelled and observed flow duration curves for Teddington natural flows

The large differences in modelled and observed natural Teddington flows below 4000 Ml/d explain the big differences between Pywr and WARMS2 modelled storages which I have commented on earlier, with an example below[3]:

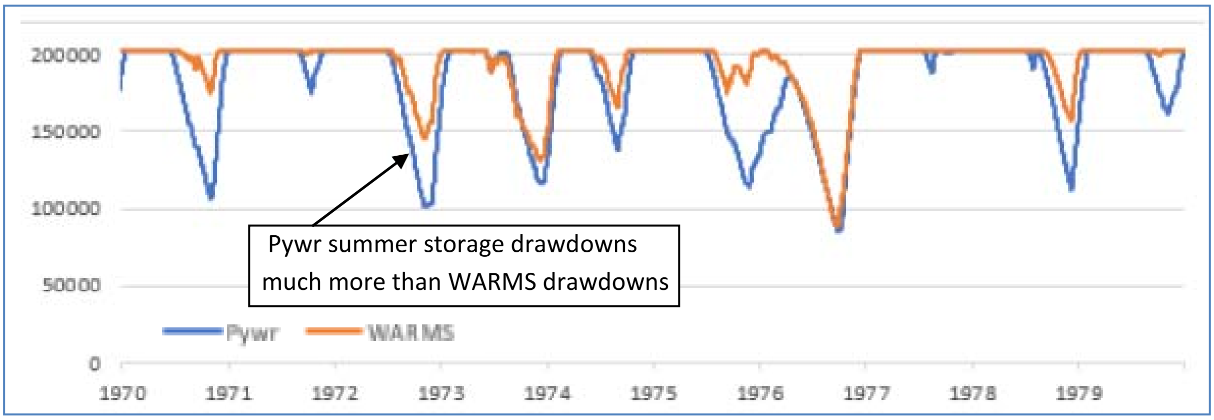


Figure 6 – Differences between Pywr and WARMS2 modelled London storages

It is hard to say whether the big discrepancies in Pywr modelling shown in Figures 5 and 6 are the result of flaws in the stochastic rainfall data or deficiencies in the rainfall-run-off modelling that converts rainfall to river flows. However the cause is immaterial – the

[3] Copied from Figure I-7 in Appendix I to TW's revised WRMP24

outcome is major flaws in the Pywr modelling.

As well as affecting the assessment of deployable output and drought resilience of Abingdon reservoir, the discrepancies in modelling of Teddington natural flow potentially has an impact on the frequency of need for the Severn-Thames transfer and hence its operating cost and carbon impact.

2.3 Validation of Pywr modelling of STT options

Appendix I to Thames Water's WRMP24 provides no details of validation of the stochastically generated river flows used in Pywr modelling of Severn-Thames transfer options or of the Pywr modelling of the transfer and its benefits to Thames Water's supplies. This is discussed further in Section 4.1 of this paper.

3. The deployable output and resilience of Abingdon reservoir

3.1 GARD's conclusions on Abingdon reservoir DO in WRMP24 response

In GARD's Addendum to the response to Thames Water's WRMP24 consultation, it was concluded that the deployable output of Abingdon reservoir has been grossly over-estimated for WRMP24 and the Gate 2 reports. In addition to failure to properly consider resilience to long duration droughts, the following flaws were identified in Thames Water's deployable output assessments:

	150 Mm^3 reservoir	**100 Mm^3 reservoir**
DO with climate change as WRMP24	271 Ml/d	185 Ml/d
Less Double counting of droughts	-6 Ml/d	-4 Ml/d
Wrong value of Culham MRF	-2 Ml/d	-1 Ml/d
Wrong climate change scenario	-19 Ml/d	-16 Ml/d
Inadequate dead & emergency storage	-44 Ml/d	-25 Ml/d
Corrected Deployable Output	**200 Ml/d**	**139 Ml/d**

GARD proposed changes to reservoir DO (excluding long drought resilience)

In addition, we consider that the deployable output of Abingdon reservoir will be a lot less than shown in the table above, perhaps only half these values, when proper consideration has been given to the likelihood of a sequence of dry years which prevent the reservoir from being full at the start of a major drought or delay its refilling after a major drought.

Thames Water's SoR rejects all GARD's findings and says on page 112 of their Appendix G2:

> *"We have not made changes to our plan as a result of this response, as our consideration is that the Deployable Output of the reservoir has been calculated robustly and in line with guidance."*

In the following sections, the evidence previously presented to support GARD's conclusions is followed by Thames Water's response.

3.2 Double counting of droughts

Extract 3 from GARD's Addendum pages 28-29

> Thames Water calculates deployable outputs for London by using the Pywr model to simulate the frequency of London reservoir storage falling into the Level 4 emergency storage zone. The London demand that causes only 38 failures in 19,200 years of simulation is then the 1:500 year deployable output (19,200 ÷ 500 = 38.4). The Pywr model is run repeatedly with small stepped increases in demand to determine the frequency of failure at each demand level and hence deployable output.

Extract 3 from GARD's Addendum pages 28-29 continued …

Under EIR-22-23-390, Thames Water provided GARD with the Pywr 'Control Line Crossing Data' showing years of failure at each modelled demand level for the existing London supplies, the 150 Mm^3 Abingdon reservoir, the unsupported 500 Ml/d STT and the STT with 500 Ml/d of support sources[21]. Using these data, we have re-assessed the deployable output of existing London supplies and the 150 Mm^3 Abingdon reservoir in Figure 14:

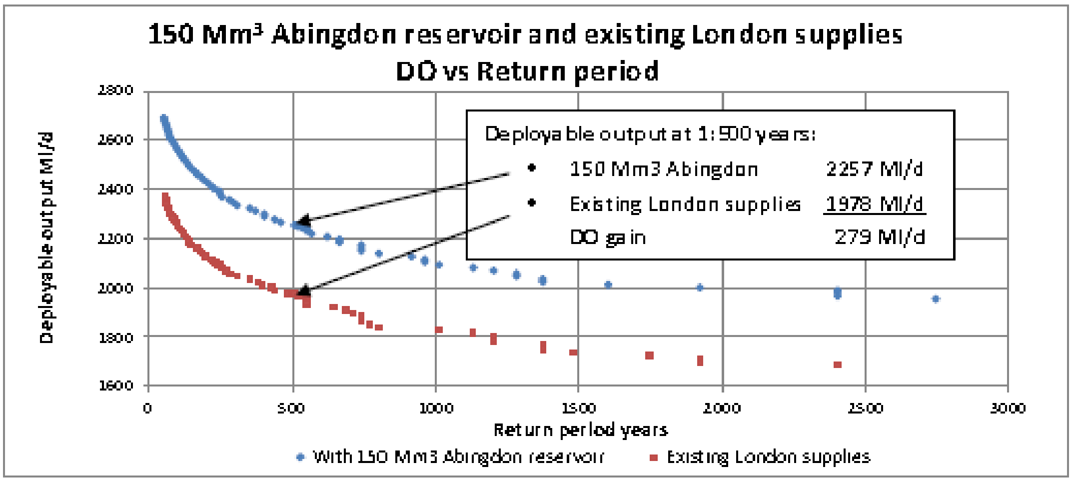


Figure 14 - GARD reassessment of Abingdon DO using Pywr model output

This shows the DO of the 150 Mm3 Abingdon reservoir to be 279 Ml/d as compared with TW's figure of 285 Ml/d. The reason for the difference is that in creating the plots in Figure 14, we have counted drought failure <u>events</u> rather than <u>years</u> of failure. Thames Water have attempted to count failure events by taking years as ending on 31st March, as explained in EIR 22-23-390[22]: *"A year is defined from Apr to Mar, in order not to count L4 events which extend into January."* Unfortunately, L4 failures in some droughts still extend beyond 31st March and some failures start before 1st April, as for the example shown below[23]:

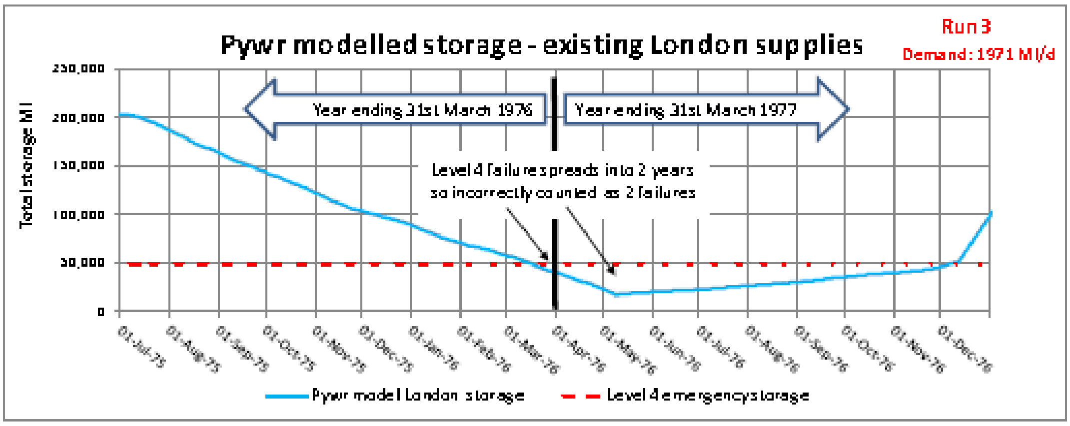


Figure 15 - Example of double counting of droughts in TW's DO analysis

The double counting of some drought events has caused Thames Water to over-estimate the deployable output of the 150 Mm^3 Abingdon reservoir by 6 Ml/d or 2%. The equivalent error for the 100 Mm^3 reservoir would be 4 Ml/d. Although not a large error, this could still be significant when comparing the reservoir with other options.

[22] EIR-22-23-390, Item 5, last paragraph
[23] Pywr model output for existing London supplies, as provided by EIR 22-23-390, Item 1

Thames Water's response to this was:

From page 112 of Appendix G2 of TW's SoR:

> *We do not agree that the approach taken in our Deployable Output calculation is incorrect. The Water Resources Planning Guideline states that we should plan so that our system is resilient to a 0.2% annual chance of failure caused by drought, where failure is defined as implementing an emergency drought order. We have, as GARD note, considered a year to be from April to March, and as such events which span across years represent additional failures.*
>
> *We note two additional factors:*
>
> - *The impacts which GARD note are very minor.*
> - *GARD have not been even-handed in their assessment, and have not considered whether the same issue should, in their consideration, impact the Deployable Output of the STT.*

I do not accept the excuse "*We have, as GARD note, considered a year to be from April to March, and as such events which span across years represent additional failures*". The methodology which considers a year to run from April to March is clearly intended to avoid counting the use of a single emergency drought order that span across a year end as two failures – as stated in by TW in their EIR 22-23-390 *"A year is defined from Apr to Mar, in order not to count L4 events which extend into January."* The fact that TW's analysis failed to notice that some emergency drought orders spanned the re-defined March-April year end is clearly an error and should be corrected, as I have done in my analysis which shows that the DO of the 150 Mm^3 reservoir is over-estimated by 6 Ml/d.

I do not agree with Thames Water's comment that the impact of this error is "*very minor*". Multiple small errors can add up to large errors so all errors should be corrected.

3.3 Use of incorrect value of Culham MRF

There was another serious Pywr modelling error in the original WRMP24 analysis in assuming that, when refilling Abingdon reservoir, the minimum required flow (MRF) in the River Thames at Culham is set at only 450 Ml/d instead of the correct value of 1450 Ml/d. TW recognised this error and provided a correction in an appendix to the modelling technical report, showing that it only reduces deployable output by 2 Ml/d. Our modelling shows a similar DO reduction due to this error, when simulating stochastic versions of the 1975-76 drought. However, the potentially slow rate of refill of Abingdon reservoir has implications for the resilience of Abingdon reservoir in multi-year droughts. GARD's concern over this is replicated in Extract 4 from the Addendum to the WRMP24 response:

Extract 4 from GARD's Addendum pages 30-31

There is another serious Pywr modelling error in assuming the minimum required flow (MRF) at Culham to be only 450 Ml/d instead of 1450 Ml/d. This error is recognised by Thames Water in a footnote[24] to the technical note on deployable assessment supplied to GARD under EIR-22-23-390 (but not available as part of the on-line WRMP24 documentation):

"When the flow is above 1,450 Ml/d then [the Pywr model assumes that] the full abstraction is available, which means that the river flow downstream of the abstraction can fall to 450 Ml/d under an abstraction of 1,000 Ml/d. This modelling of the constraint as a simple threshold was identified as incorrect towards the end of the study, and a Minimum Residual Flow (MRF) of 1,450 Ml/d should have been applied instead (so that at a river flow of 1,500 Ml/d then only 50 Ml/d can be abstracted). The impact of the MRF constraint is described in Appendix B."

Appendix B of the technical note shows that this Pywr modelling error has no modelled effect on the deployable output of the 150 Mm^3 reservoir using river flows without climate change and reduces the deployable output with climate change by 2 Ml/d (reducing the deployable output from 271 Ml/d to 269 Ml/d[25]). GARD's modelling using stochastic flows without climate change also shows the error does not affect deployable output. We do not have the climate change flow data to check the 2 Ml/d DO loss.

Although the Culham MRF error does not have a big impact on deployable output, it can greatly affect the speed of reservoir refilling after droughts. The main Gate 2 report for Abingdon reservoir claims that the reservoir refills in 5 months after droughts:

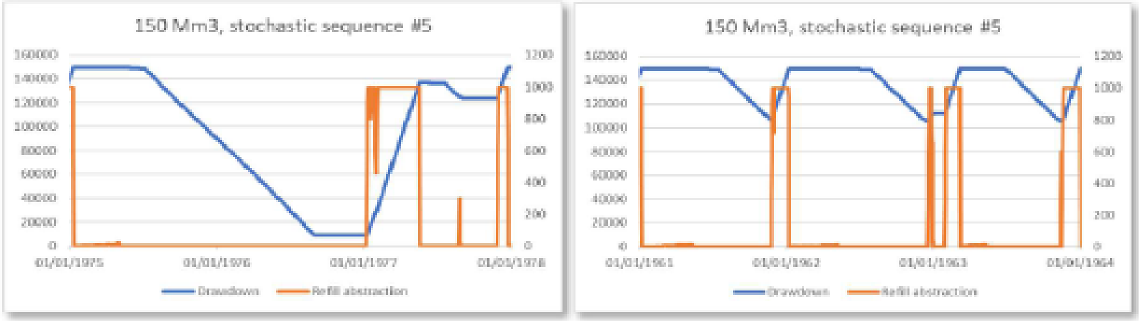
"Following drought periods, which result in longer periods of reservoir release to meet demands for water and hence a lower and deeper drawdown period, abstraction refill occurs for longer during the subsequent refill season as greater volumes are required to refill the reservoir. However, even after a long period of extreme drought and drawdown, refill is still achieved within 5 months. This is illustrated in Figure 4.1 below for one of the synthetic stochastic hydrological sequences. Refill would tend to be faster for the smaller reservoir sizes, due to the reduced volumes of storage."

Thames Water's Figure 4.1 which is misleadingly said to illustrate 5-month refill after an extreme drought is copied below:

[24] Technical Note Enhanced RSS Modelling of SESRO and Thames to Affinity Transfer Schemes, footnote 2, page 3
[25] Ibid, Appendix B, Table 6-7

Extract 4 from GARD's Addendum pages 30-31

Figure 4.1 SESRO drawdown and refill – extreme drought (left) compared to standard operation (rig



** Note, primary y-axis is drawdown (Ml) and secondary y-axis is refill abstraction (Ml/d)*

Figure 16 - TW illustration of rapid Abingdon refill after "extreme" drought

The rapid refill shown in the Gate 2 report's Figure 4.1 will have been enhanced by the incorrect assumption of a 450 Ml/d MRF at Culham. It should also be noted that the historic drought of 1976 was followed by a wet winter, so the numerous stochastic droughts simulated by perturbation of historic 1976 weather also tend to be followed by wet winters and rapid Abingdon reservoir refill. When the stochastic droughts are based on the historic drought of 1992, which was followed by another quite dry year, Abingdon reservoir usually fails to refill in the next year, as for the example below generated by GARD modelling using Run 86 stochastic data and the correct 1450 Ml/d MRF at Culham:

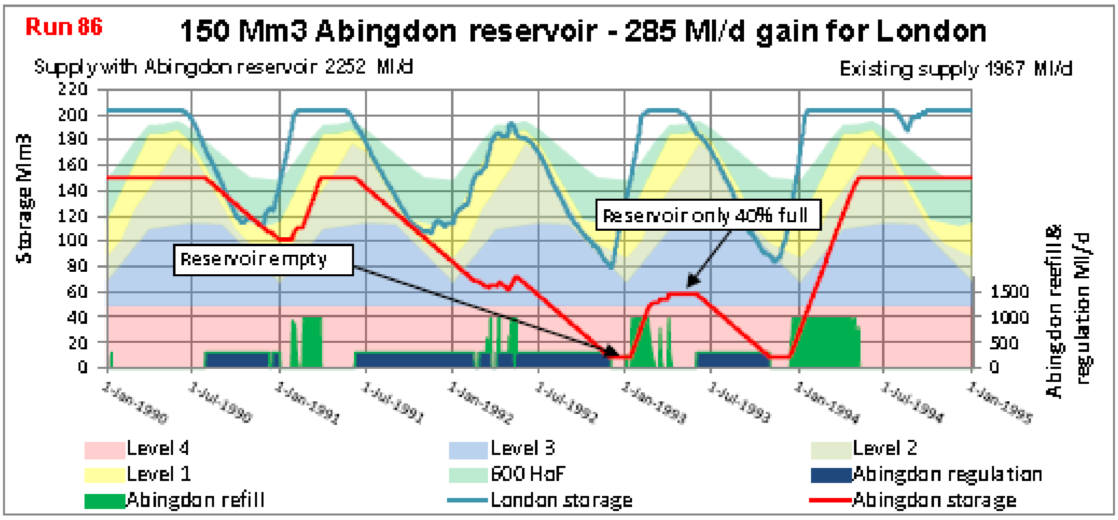


Figure 17 - Example of Abingdon reservoir failing to refill in the year following a drought

The 1991 drought in Run 86 was the 48th most severe drought in the 19,200 year stochastic record, so it has a return period of 1:400 years. Although London storage does not fall to Level 4 in this drought (because it is less severe than 1:500 years), Abingdon reservoir is still empty by the end of the drought and only 40% full at the start of the drought of 1993, quickly dropping again to empty. This is a good example of the reservoir's poor resilience in long droughts, which we will consider more in Section 3.3.

Thames Water's response to the criticism of using the incorrect value of the Culham MRF is shown below:

From pages 112-113 of Appendix G2 of TW's SoR:

The error noted around the representation of the Culham MRF was fixed in the Deployable Output modelling prior to publication of rdWRMP24, and as GARD note the Deployable Output impact was very minor (2 Ml/d).

We agree that a HOF [hands-off flow, ie minimum required flow, MRF] of 1450 Ml/d rather than 450 Ml/d in the modelling reduces the speed of refill after drought events. However, our consideration should be the Deployable Output benefit of the reservoir to our supplies, and this is what our Deployable Output calculations are designed to do. GARD's comments around the potential risks during events which may occur are hypothetical, and do not consider the resilience/vulnerability of our existing system. SESRO is particularly effective in drought events to which the London supply system is particularly vulnerable, and which GARD state that the London system is particularly vulnerable to (see p.38 of the main GARD representation), "two dry summers and an intervening dry winter".

We have not made changes to our plan as a result of this response, as our consideration is that the Deployable Output of the reservoir has been calculated robustly and in line with guidance.

As I commented on Thames Water's dismissal of their *"very minor"* 6 Ml/d drought double counting error, multiple small errors can add up to large errors, so the Culham MRF error should be corrected in the assessment of Abingdon reservoir DO.

TW recognise that correction of the Culham MRF error reduces the speed of reservoir refill after drought events. However, they have then ducked proper assessment of its implications in terms of resilience of London supplies in long droughts and the potentially prolonged supply restrictions when Abingdon reservoir fails to refill after a severe drought. I will say more about this in Section 3.6 of this paper.

3.4 Use of wrong climate change scenario in assessing reservoir DO

In the main WRMP report, the widely quoted deployable outputs for Abingdon reservoir are 271 Ml/d for the 150 Mm^3 reservoir and 185 Ml/d for the 100 Mm^3 reservoir. These are TW's assessments for the 'median' climate change scenario. However, TW's preferred plan assumes the 'high' climate change scenario, so in my opinion the assessed DOs for Abingdon reservoir should also be for the 'high' climate change scenario, not the 'median' scenario. This would reduce the deployable output of the 150 Mm^3 Abingdon reservoir by 19 Ml/d. The evidence for this is shown in Extract 5 from GARD's Addendum to the WRMP24 consultation response:

Extract 5 from GARD's Addendum pages 32-33

3.2 Effect of climate change on reservoir deployable output

The technical note on modelling of Abingdon reservoir shows the reduction in reservoir deployable output with median climate change as below[26]:

SESRO Size Mm^3	Previous DO (no CC) (Ml/d)[11]	Derived latest DO (no CC) (Ml/d)[12]	Derived CC impact (Ml/d)[13]	Derived latest DO with CC (Ml/d)
30	67.9	66.9	-1.4	65.5
75	155.0	154.2	-5.0	149.2
80	163.0	160.5	-5.4	155.1
100	195.0	192.0	-7.4	184.6
122	238.0	234.3	-10.3	224.0
125	243.9	240.2	-10.7	229.5
130	253.7	250.1	-11.4	238.6
150	292.9	285.4	-14.4	271.0

Table 2 - TW assessment of median climate change impact on reservoir DO

The median climate change deployable outputs of 271 Ml/d for the 150 Mm^3 reservoir and 185 Ml/d for the 100 Mm^3 reservoir are the figures used in Thames Water's main WRMP. The justification for assuming the median climate change impact is said to be[27]:

> *"Tier 1 DO calculation undertaken using WRSE Pywr model, involving a 'full stochastic' DO assessment, and incorporating the impact of climate change as per the WRSE standard approach to climate change assessment"*

However, Thames Water's preferred plan is based on 'pathway 4' which includes 'high' climate change[28]. Therefore, in developing the preferred plan, the deployable output for Abingdon reservoir should also allow for the high climate change scenario and not the median scenario. The effect of different climate change scenarios on reservoir deployable output is shown in the technical note as below[29]:

[26] Ibid, Table 6-4
[27] TW Main WRMP Report, Table 7-6 commentary
[28] TW Main WRMP Report, paragraphs 11.11 and 11.14
[29] Technical Note Enhanced RSS Modelling of SESRO and Thames to Affinity Transfer Schemes, Table 6-1

Extract 5 from GARD's Addendum pages 32-33 continued …

SESRO option 75 Mm³		SESRO option 150Mm³		Comment
Climate change scenario	Net DO impact (Ml/d)	Climate change scenario	Net DO impact (Ml/d)	
cc_01	-16.0	cc_01	-42.8	Worst-case
cc_03	-15.8	cc_03	-40.7	
cc_26	-14.0	cc_02	-36.5	
cc_02	-13.6	cc_06	-34.9	
cc_13	-12.8	cc_13	-34.6	
cc_11	-12.4	cc_10	-31.0	
cc_10	-11.5	cc_26	-30.5	
cc_06	-11.2	cc_11	-29.5	
cc_14	-9.0	cc_14	-26.4	
cc_05	-7.3	cc_05	-24.1	
cc_18	-6.9	cc_09	-20.7	
cc_24	-5.9	cc_04	-16.2	
cc_09	-5.3	cc_24	-16.0	
cc_04	-5.1	cc_17	-15.6	Median (lower of average)[10]

Table 3 - Effect of climate change scenarios on 1:500 year reservoir DO

Thames Water's WRMP appendix on climate change says that scenario cc_06 is used as the 'high' climate change scenario[30]. Table 3 shows that reduction in deployable output for the high scenario should have been 34.6 Ml/d for the 150 Mm³ reservoir rather than 15.6 Ml/d reduction that was assumed for the median scenario.

In other words, the deployable output of the 150 Mm³ reservoir with climate change allowance should have been 252 Ml/d, not 271 Ml/d. The equivalent deployable output of the 100 Mm³ reservoir with climate change should have been 169 Ml/d, not 185 Ml/d.

Combined with the correction of the errors due to double counting of droughts, TW's assessed deployable outputs reduce to 246 Ml/d for the 150 Mm³ reservoir and 165 Ml/d for the 100 Mm³ reservoir. These figures are before consideration of the resilience of the reservoir to long duration droughts or the adequacy of the allowance for emergency storage.

[30] TW WRMP Appendix U Climate Change, paragraphs U.77 and U.101

Thames Water's response to the criticism over use of the wrong climate change scenario in assessing DO is shown below:

From pages 112-113 of Appendix G2 of TW's SoR:

> *We have considered a median climate change impact for both the SESRO and STT options, and our consideration is that applying a median climate change reduction is appropriate and even-handed. We do not agree that we should adopt the "High" climate change scenario in the assessment of option Deployable Output calculation, as the Deployable Output figures are used in all branches of our adaptive plan. Applying a climate change impact reduction to option Deployable Output values is a step taken to ensure that we have considered climate impacts in our option assessment, and our primary concern in this respect is ensuring that we are comparing alternatives against one another in an appropriate way.*
>
> *We note that GARD have again not taken an even-handed approach in their consideration of which factors to consider in the calculation of different options' Deployable Outputs.*
>
> *We have not made changes to our dWRMP following this response, for the reasons set out in our consideration.*

In my opinion, Thames Water's arguments are irrational and unacceptable. The use of the of median climate change scenario in assessing DO for all branches of the adaptive plan is an error which should be corrected. Obviously, if Thames Water is comparing options for different climate change scenarios, the deployable outputs of options need to be adjusted for each climate change scenario. Otherwise, options that are climate change resilient like leakage reduction, metering, re-cycling and desalination will be unfairly assessed in comparison with options whose DOs reduce with climate change, like reservoirs.

In the case of Severn to Thames transfer options, I agree that the DOs should be adjusted for each climate change scenario, but note that several of the STT support options are climate change resilient like the Netheridge and Minworth recycling options.

3.5 Inadequate allowances for dead and emergency storage

GARD's WRMP response proposed that TW's planned 6% emergency storage allowance for Abingdon reservoir should be increased to 15% to be in line with the emergency storage allowance in other major UK reservoirs. GARD also proposed that, for the water in emergency storage to be of sufficiently good water quality to be useable, dead water should be based on an average residual water depth of 5m, not an average depth of 2.5m as planned by TW. With these proposals for dead storage and emergency storage, GARD's modelling shows that the deployable outputs for the 150 Mm^3 and 100 Mm^3 reservoir would reduce by 44 Ml/d and 25 Ml/d respectively.

GARD's evidence for this is copied below in Extract 6:

Extract 6 from GARD's Addendum pages 37-41

3.4 Allowances for dead and emergency storage

In our main response to the consultation on Thames Water's WRMP24, we proposed that TW's proposed 6% emergency storage allowances for Abingdon reservoir should be increased to be in line with the emergency storage allowance in other major UK reservoirs as below:

- Clywedog reservoir 13%
- Llyn Brianne reservoir 14%
- Bristol Water (Chew, Blagdon) 18%
- Welsh Dee system 20%
- TW London reservoirs 24%
- TW Farmoor reservoir 33%

Thames Water says that the allowance of 6% emergency storage, ie 9,000 Ml for the 150 Mm^3 reservoir, is equivalent to 30 days of supply from the regulation release of 300 Ml/d, which they claim is in line with UK normal practice. However, there appears to have been no consideration of the minimum average depth of water required for acceptable water quality. Thames Water's themselves agree that an average water depth of less than 5m will be likely to lead to water quality problems[35]:

> *"The 28m water depth noted in the [GARD's] comment is the depth of the live storage (51m AOD to 79m AOD), there is a further 5m depth of dead storage in the central trench underneath (46m AOD to 51m AOD). We agree that a water depth of less than 5m would likely lead to water quality issues, hence the definition of such water as dead storage."*

Therefore there should be a minimum average depth of 5m of water when the emergency storage is <u>empty</u>. Figure 21 shows a cross-section of the reservoir and borrow pit[36]:

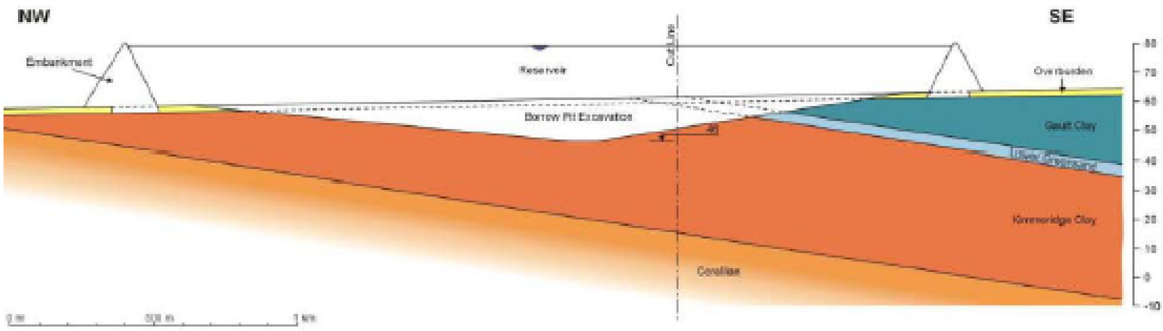


Figure 21 - Cross-section of reservoir showing borrow pit

This shows that the maximum depth of the borrow pit is about 5m so the average depth is only about 2.5m, not 5m. The average depths of water for the dead storage and Thames Water's proposed emergency storage are shown in Table 4:

---

35 WRMP19 Reservoir Feasibility Report, page 435, Mott MacDonald, July 2017

36 Gate 2 Concept Design Report Figure 2.1

Extract 6 from GARD's Addendum pages 37-41 continued …

Reservoir dimensions from 2017 reservoir feasibility report[37]	150 Mm^3 reservoir	100 Mm^3 reservoir
Gross storage	165,000 Ml	110,000 Ml
Live storage	150,000 Ml	100,000 Ml
Dead storage	15,000 Ml	10,000 Ml
TW emergency storage (6% of live storage)	9,000 Ml	6,000 Ml
Area at full supply	675 ha	404 ha
Embankment perimeter	10.3 km	7.9 km
Area at base of embankment	551 ha	309 ha
Average depth of dead storage	2.72 m	3.23 m
Maximum depth of TW emergency storage	1.63 m	1.94 m
Average depth, dead + maximum emergency	**4.35 m**	**5.17 m**

Table 4 - TW proposed water depths for dead and emergency storage

This shows that Thames Water's planned volumes of dead and emergency storage fail to meet their own criterion for a minimum average depth of 5m for useable water. None of Thames Water's proposed emergency storage for the 150 Mm^3 reservoir would be useable because it would all have to come from an average water depth of less than 5m. Only 0.17m depth of the proposed 6,000 Ml of emergency storage for the 100 Mm^3 reservoir would be useable, equivalent to just 525 Ml.

Thames Water's emergency storage proposals ignore their own concerns about future water quality as stated in the main WRMP24 report[38]:

> *"By looking at the resilience of our raw water storage and supply network we have found that the change in algal bloom severity and duration is dependent on individual reservoir characteristics, including their physical structure and management. For example, deeper reservoirs have better control measures to manage the raw water quality and therefore are more resilient to the impacts of climate change.*
>
> *Nevertheless, as well as future raw water resource availability, the water quality challenge and how this may change in future climates is an important factor to account for in planning. Evidence indicates that the impact of climate change is increasing the range of species of algae that can cause a bloom event in our reservoirs and also increasing the period of year for which our reservoirs are at risk of algal bloom."*

Recognising the increasing threat of algal blooms and poor reservoir water quality, we propose that the allowances for dead and emergency storage should be:

- Dead water should be based on an average residual water depth of 5m
- Emergency storage should be 15% of live storage to be in line with Llyn Brianne, Clywedog and the Welsh Dee regulating reservoirs

[37] WRMP19 Reservoir Feasibility Report, PDF pages 242-243 and 248-249
[38] TW WRMP24 main report, paragraphs 4.129 and 4.130

Extract 6 from GARD's Addendum pages 37-41 continued …

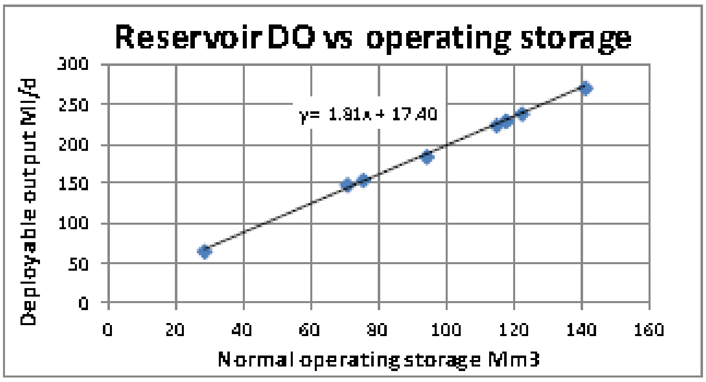
In our opinion, these would be reasonably cautious allowances to make, in line with the precautionary water quality measures being adopted for the STT, including the treatment of all transferred water at Deerhurst and high levels of treatment planned for Minworth and Netheridge effluent.

The reassessed dead and emergency storage volumes would then be as below:

GARD reassessment of dead and emergency storage	Nominal capacity: 150 Mm^3 reservoir	Nominal capacity: 100 Mm^3 reservoir	Comment
Gross storage	165,000 Ml	110,000 Ml	As per 2017 feasibility report
Dead storage with average 5m depth	27,570 Ml	15,460 Ml	Bottom area ha x 5m depth
Live storage, including emergency	137,430 Ml	94,540 Ml	Gross storage less dead
Emergency storage 15% of live storage	20,615 Ml	14,181 Ml	15% typical for regulating reservoirs
Storage available for normal operation	**116,816 Ml**	**80,359 Ml**	Live storage less emergency
Average depth of dead storage	5.0 m	5.0 m	TW stated minimum acceptable
Average depth of GARD emergency storage	3.7 m	4.6 m	Emergency storage ÷ bottom area
Average depth dead + emergency	**8.7 m**	**9.6 m**	Depth remaining at start of emergency

Table 5 - GARD reassessment of dead and emergency storage allowances

The relationship between normal operating storage and reservoir deployable output assuming median climate change is as below, using the same data from the SESRO modelling technical note as used in our Table 2 [39]:



Storage: Nominal Mm3	Storage: Normal operation Mm3	DO Ml/d
30	28.2	65.5
75	70.5	149.2
80	75.2	155.1
100	94.0	184.6
122	114.7	224.0
125	117.5	229.5
130	122.2	238.6
150	141.0	271.0

Figure 22 - Abingdon reservoir deployable output vs normal operating storage

Combining the changes in normal operating storage shown in Table 5 with the trendline relationship between storage and DO shown in Figure 22, the impact on Abingdon reservoir DO of GARD's proposals for dead and emergency storage is shown in Table 6:

[39] Technical Note Enhanced RSS Modelling of SESRO and Thames to Affinity Transfer Schemes, Table 6-1

Extract 6 from GARD's Addendum pages 37-41 continued …

	Normal operating storage			
Option	**TW**	**GARD**	**Difference**	**DO reduction**
150 Mm³ reservoir	141,000 Ml	116,816 Ml	24,185 Ml	43.8 Ml/d
100 Mm³ reservoir	94,000 Ml	80,359 Ml	13,641 Ml	24.7 Ml/d

Table 6 - Reservoir DO reduction with GARD proposed dead and emergency storage

With GARD's proposals for dead storage and emergency storage, Table 6 shows that the deployable outputs for the 150 Mm^3 and 100 Mm^3 reservoir would reduce by 43.8 Ml/d and 24.7 Ml/d respectively.

Thames Water's response to the criticism of the inadequacy of dead and emergency storage is shown below:

From pages 112-113 of Appendix G2 of TW's SoR:

We do not agree with the amendments which GARD suggest to the dead/emergency storage provisions for SESRO, for the reasons set out below. As such, we do not agree the Deployable Output reductions suggested.

The 6% emergency storage in SESRO is calculated as 30 days' worth of reservoir throughput, in line with other Thames water reservoir emergency storage calculations. Given that this is the standard on which other TW emergency storage requirements are determined, in the absence of other evidence we do not see a reason to amend this.

The reservoir water quality modelling that was undertaken for the Gate 2 submission suggests that an acceptable level of water quality can be achieved with the current concept design and associated inlet / outlet and mixing arrangements. This will continue to be reviewed and re-analysed as the design progresses, to reflect the latest design of the reservoir and borrow pit, and appropriate inlet, outlet and mixing arrangements included as required.

Our consideration is that GARD have taken the suggestion of 5m being required to ensure good water quality out of context. GARD have assumed that 5m depth is required, on average, to ensure good water quality, when this was intended to apply to the depth of storage required in the central trench to ensure good quality, considering the rest of the design of the reservoir (i.e., including the sloping banks of the borrow bit) and accounting for the aeration system which encourages mixing.

Thames Water's only justification for the 6% of emergency storage is that it is *"30 days' worth of reservoir throughput, in line with other Thames water reservoir emergency storage calculations".* **This statement is wrong** – the emergency storage allowance in Thames Water's existing reservoirs is actually about 70 days of "throughput".

The throughput of the London reservoirs in droughts is about 760 Ml/d, as seen in Thames Water's Aquator modelling of the London supply system at a deployable output of 2302 Ml/d as shown below[4]:

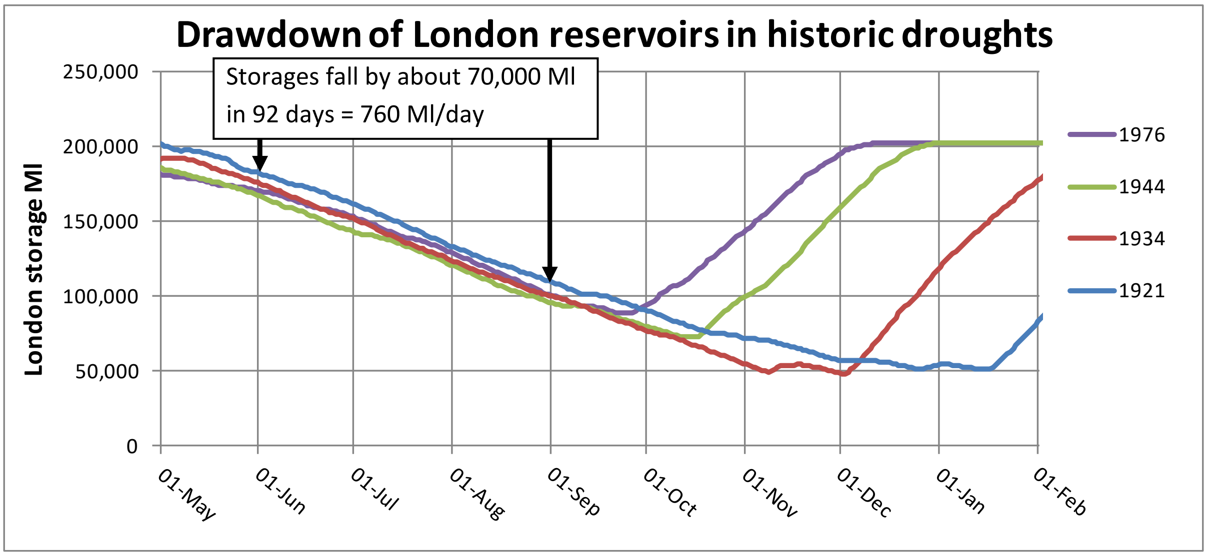


Figure 7 - TW modelled throughput of London reservoirs in severe droughts

The emergency storage allowance for the London reservoirs is 48,500 Ml which is 64 days of throughput, not 30 days. There is a similar picture for Farmoor reservoir[5]:

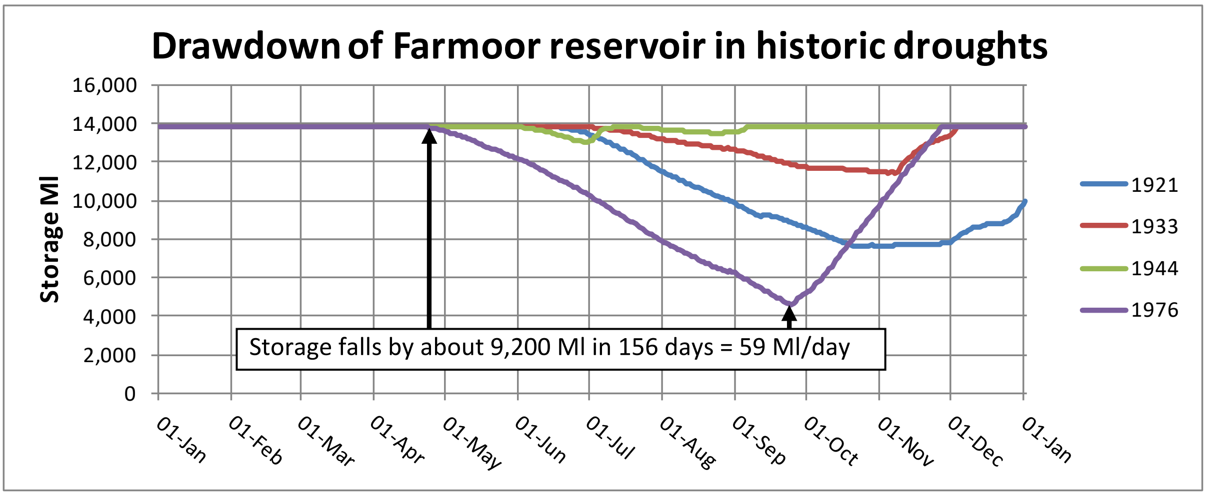


Figure 8 - TW modelled throughput of Farmoor reservoir in severe droughts

The emergency storage for Farmoor is 4500 Ml, equivalent to 76 days of 59 Ml/d throughput.

[4] Data from EIR-22-23-390 file *"AR20 Q4 Sc1 - London DO – 2302.xlsx"*
[5] Data from WARMS2 output supplied to GARD in file *"GARD Upper Thames DO 245 AR16.xlsx"*

Thames Water's allowance of 6% emergency storage in Abingdon reservoir is far lower than allowed in other UK reservoirs, as shown below with references for the source data:

- Clywedog reservoir 13%[6]
- Llyn Brianne reservoir 14%[7]
- Bristol Water (Chew, Blagdon) 18%[8]
- Welsh Dee system 20%[9] (of gross storage)
- TW London reservoirs 24%[10]
- TW Farmoor reservoir 33%

Yorkshire Water's policy is for *"30 days supply at the reservoir or group yield, or 12.5 per cent of reservoir stocks, whichever is greater"*. [11]

Bearing in mind the poor winter refill of Abingdon reservoir and the vulnerability of the London supply system to longer duration droughts, GARD's proposal that the Abingdon reservoir emergency storage should be 15% of live storage seems reasonable. For the 137,400 Ml live storage that GARD has calculated for the nominal 150 Mm^3 reservoir, 15% emergency storage would be 20,600 Ml which is 64 days of throughput of the planned regulation release of 321 Ml/d – consistent with the 64 days of throughput for emergency storage in Thames Water's London reservoirs.

Thames Water's justification of an average dead storage water depth of only 2.5 m (ie 5 m maximum depth at the deepest point of the borrow pit) seems to be dependent on water quality modelling which I have not seen. Bearing in mind the incorrect analysis that Thames Water has used to justify their 6% Abingdon reservoir emergency storage, I suggest the dead storage allowance needs to be carefully scrutinised by the Environment Agency, taking account of experience of algal blooms in other reservoirs and the impact of releasing algae-laden water into the River Thames.

3.6 Resilience of Abingdon reservoir to long duration droughts

GARD' Addendum to the response to Thames Water's WRMP24 consultation concluded that, if proper consideration is given to the occurrence of long duration droughts, the deployable output of Abingdon reservoir would be far less than that claimed by Thames Water, perhaps in the region of only 50% of the claimed amounts. GARD's evidence for this is shown in Extract 7 from the Addendum:

---

[6] South Staffs Water Draft Drought plan, Figure 2, August 2017
[7] DCWW Welsh Water Drought Plan, Figure 20, July 2015
[8] Bristol Water Draft Final Drought Plan, Figure 2, July 2017
[9] United Utilities Revised Draft Drought Plan, Figure A6.11, January 2017
[10] TW WARMS2 modelling of London and Farmoor systems, as provided to GARD
[11] Yorkshire Water Draft Drought Plan, Section 2.1, January 2018

Extract 7 from GARD's Addendum pages 33-36

3.3 Reservoir resilience in long duration droughts

In the run-up to WRMP19 and in our responses to the two consultations on WRMP19 GARD demonstrated the lack of resilience of Abingdon reservoir in multi-year droughts and the flaws in Thames Water's method of assessment of resilience in long droughts[31] [32]. The stochastic records used for assessment of schemes in WRMP19 included many droughts in which the 150 Mm^3 Abingdon reservoir would be unable to deliver its supposed deployable output of about 290 Ml/d. An example of such a drought is shown below:

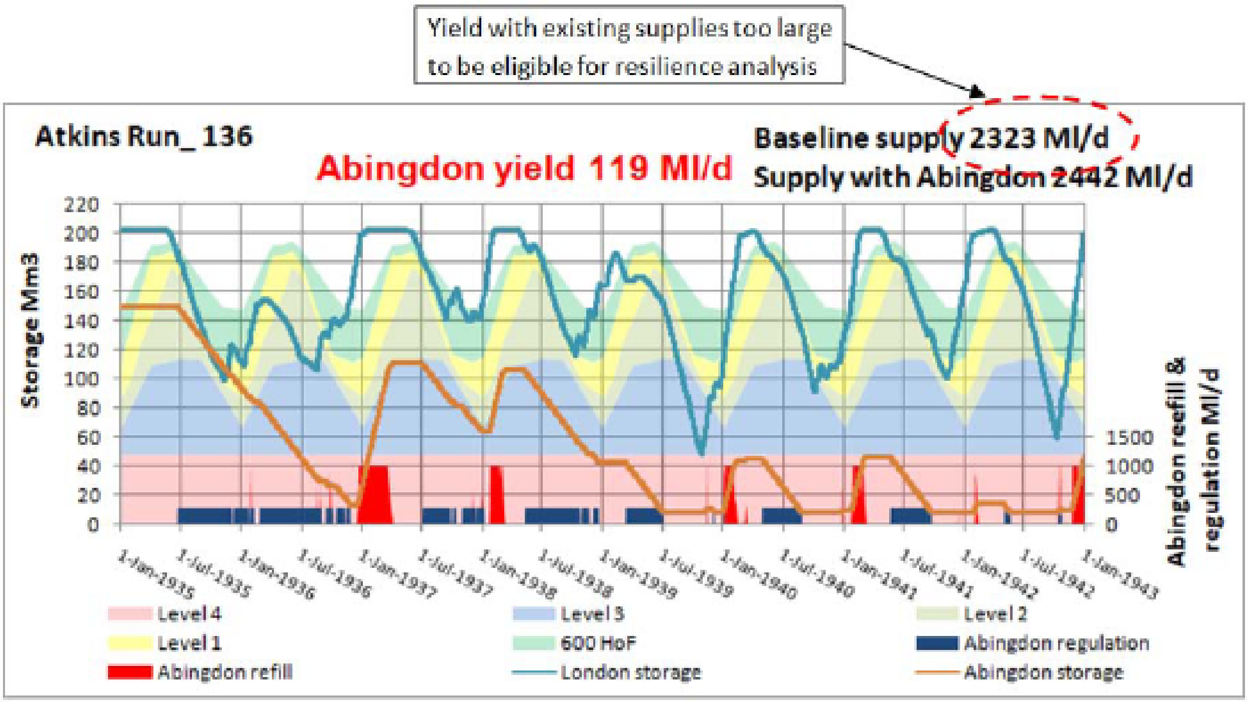


Figure 18 - Example of catastrophic drought not considered by TW in WRMP19

In our response to WRMP19, we commented[33] that, for the example above, the existing London supplies would be only moderately tested in the drought of 1939 (return period of less than 1:100 years for existing London supplies), so the drought was not sufficiently severe to be selected by TW for resilience checking. However, the succession of dry winters leading up to 1939 would leave Abingdon reservoir only about 30% full at the start of the 1939 drought. After catastrophic failure of London's supplies in 1939, Abingdon reservoir would remain virtually empty for several years, leaving London's supplies in a state of prolonged crisis. This type of event would be a major risk for London because of the severity of its consequences, but Thames Water's WRMP19 methodology failed to identify it or provide an estimate of its probability.

[31] GARD response to first WRMP consultation, pages 134-141, April 2018

[32] GARD response to second WRMP consultation, pages 111 to 117, November 2018

[33] GARD response to 2nd consultation on WRMP19, Figure 8-2 and last paragraph on page 115 https://www.abingdonreservoir.org.uk/downloads/GARD%20%20response%20to%202nd%20Consultation%20on%20TW%20draft%20WRMP%20Rev%2029.11.18.pdf

Extract 7 from GARD's Addendum pages 33-36 continued …

To some extent, the flaws in the WRMP19 method of assessment have been addressed by the Pywr modelling of the full 19,200 years of stochastic data, rather than the previous method which selected only a small proportion of droughts in the stochastic record, excluding many long duration droughts like the example shown in Figure 18.

However, as discussed in Section 3.1, the new method of generating stochastic data, excluding the long droughts of 1933/34 and 1943/44 from the base 'training' period, has introduced a new bias whereby most of the droughts in the 19,200 year record are based on the shape and duration of the 1975/76 drought which was not particularly long and was not preceded or followed by dry years. The dominance of the 1975/76 drought in shaping the droughts in the new historic record is shown in Figure 7 of this Addendum.

Nevertheless, even using the new flawed stochastic data, with their lack of long droughts, the modelling raises doubts about the resilience of the reservoir in the relatively few long droughts in the stochastic record. Figure 18 shows GARD modelling of the performance of the reservoir delivering the expected 285 Ml/d deployable output (no climate change) in the 1954 drought of Run 218. This drought is the 38th most severe drought in the drought sequences shown in Figure 2, so it has a return period of 1 in 505 years:

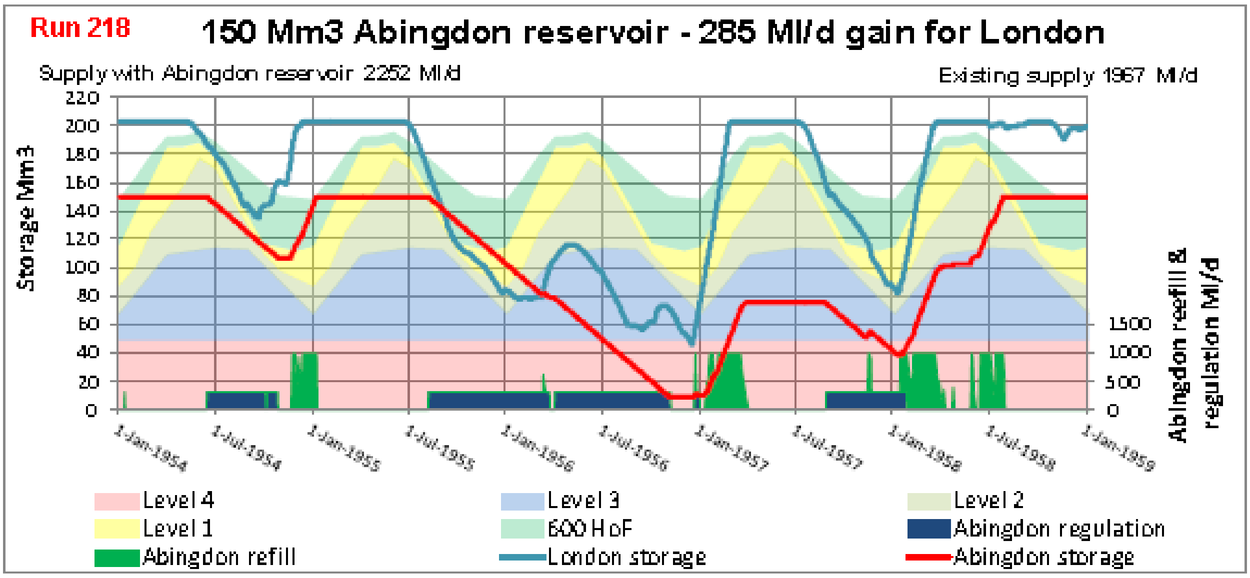


Figure 19 - Operation of 150 Mm3 Abingdon reservoir in a 1:505 year drought

This simulation illustrates some of the concerns over resilience of Abingdon reservoir in long droughts:

1. The simulated drought of 1955 was moderately severe and half the Abingdon storage had been used by the end of the summer. There would have been minimal refill during the winter of 1955-56, so Abingdon reservoir and the London reservoir would still be half full at the start summer 1956, with London already subject to Level 3 restrictions. These restrictions would remain in place for virtually the whole year. This would seem an unacceptable level of service for London.

Extract 7 from GARD's Addendum pages 33-36 continued …

2. Abingdon reservoir falls to the emergency storage level of 9,000 Ml on 5th November 1956, whereas the London storage continues to fall for about 6 weeks reaching its minimum level on 16th December. Thames Water have not said in their Gate 2 reports whether reaching the emergency storage level in Abingdon reservoir would trigger Level 4 emergency measures in London. Even if it doesn't, with London reservoirs nearing the emergency level and Abingdon reservoir effectively empty, there would surely need to be restrictions in London demands beyond the Level 3 measures which have relatively little effect in the autumn months.
3. At the start of the following summer, Abingdon reservoir would still have been less than half full. If the drought in the late summer of the second year started earlier than July, there would have been a major problem with extended Level 4 failure of London's supplies.

The fundamental problem with the resilience of Abingdon reservoir in long droughts is that there is minimal water available to refill it in even moderately dry winters. Therefore, it is vulnerable to a succession of 3 or more dry years.

With the method that has been used to generate the 19,200 years of stochastic data, the pattern of drought occurrence and the shape of individual droughts is governed by the sequence of occurrence of droughts in the historic weather period used to "train" the generated stochastic data. The historic period used, 1950 to 1997, did not contain either a drought which severely tested London's supplies (the 1976 drought was too short) or any extended sequence of drought years. Therefore, the stochastic data generated could not include any multi-year drought sequences of the type that Abingdon reservoir is unable to deal with.

As previously mentioned in Section 2.2, this flaw in the stochastic data was identified in WRSE's method statement on the stochastic climate datasets[34]:

> *"There is a risk that extreme, extended droughts may not necessarily be well reflected in the dataset, although quantifying this risk is extremely difficult. Companies may complement the stochastic dataset with drought artificial weather series to represent prolonged drought events (which the stochastic generator will not have been trained on).*

Despite this advice and the known concerns over long droughts, Thames Water has failed to consider any artificial weather series to represent prolonged drought events. By re-ordering the sequence of dry years in the historic record, it can be shown that Abingdon reservoir would fail to deliver its expected deployable output in a succession of dry years preceding a major drought.

[34] WRSE Method Statement on Stochastic Climate Datasets: Consultation Version, July 2020, paragraph 2.7

Extract 7 from GARD's Addendum pages 33-36 continued …

For example, if the historic drought of July 1933 to November 1934 had been preceded by the historic river flows of July 1996 to June 1997, the effect on Abingdon reservoir trying to deliver its expected 285 Ml/d deployable output (without climate change) would be as shown by GARD's modelling in Figure 18:

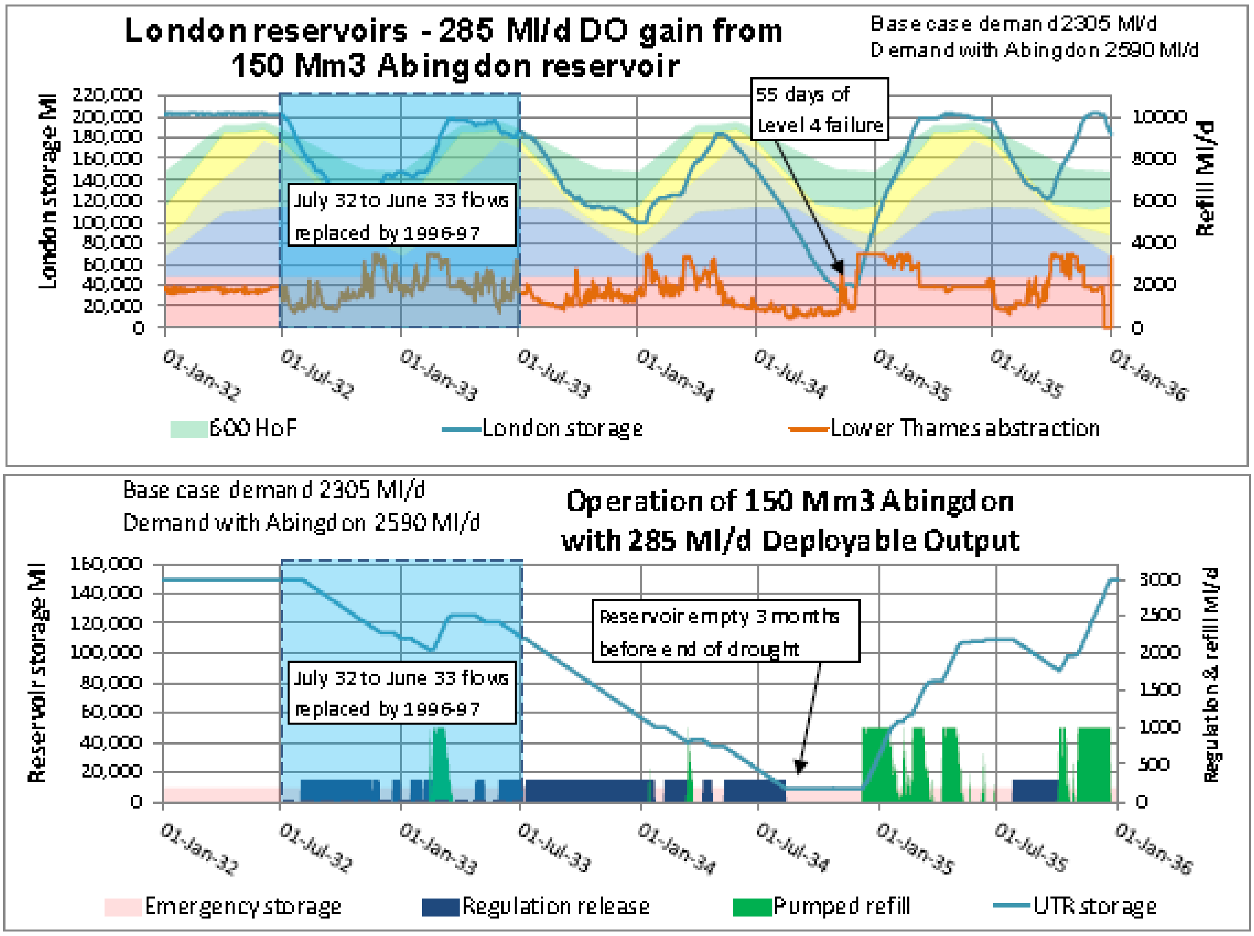


Figure 20 - Abingdon reservoir in artificially extended historic 1934 drought

In this scenario, replacing the historic flows of mid-1932 to mid-1933 with the historic flows of mid-1996 to mid-1997 would lead to 55 days of Level 4 failures for London's supplies, with Abingdon reservoir being empty 3 months before the end of the drought. This would be a catastrophic failure of London's supplies, with Level 4 restrictions starting in August 1934 at the peak of the tourism season. In this seemingly plausible scenario, the deployable output that can be sustained by Abingdon reservoir is only 163 Ml/d, not 285 Ml/d.

We conclude that, if proper consideration is given to the occurrence of long duration droughts, the deployable output of Abingdon reservoir would be far less than that claimed by Thames Water, perhaps in the region of only 50% of the claimed amounts. In the 5 years since WRMP19, Thames Water have failed to address the concerns previously raised by GARD, even after their validity had been acknowledged by WRSE in their method statement on generating stochastic climate datasets in 2020.

Thames Water's response to the criticism of their failure to consider long duration droughts starts by claiming that they have followed the Water Resource Planning Guideline and criticising GARD's general approach:

From pages 114-115 of Appendix G2 of TW's SoR:

> *The approach GARD have taken to assessing the Deployable Output benefit of the reservoir is incorrect and does not comply with the Water Resources Planning Guideline. We do not agree that we should amend our Deployable Output assessment of the reservoir to focus on long-duration droughts, for the reasons set out below.*
>
> *The Water Resources Planning Guideline states clearly (Section 5.1) that, in our calculation of Deployable Output, we should use a 'system response' approach, viz., when assessing our Deployable Output, or the Deployable Output benefit that new interventions may bring, we should consider how our different sources work together to provide resilience to drought. This guidance reflects the fact that some systems are particularly vulnerable to "short, sharp" drought events, while other systems are vulnerable to more extended but less intense drought events and thus assessment of supply capability should reflect the vulnerabilities of a given supply system. It follows that a given intervention will provide different benefits within different supply systems, according to the existing vulnerabilities of that supply system.*
>
> *GARD's focus solely on the Deployable Output "of the reservoir" is, as such, irrelevant, and it is the "Deployable Output benefit that the reservoir brings to our supplies" which we should consider. As GARD highlight in an earlier section of their response, the critical vulnerability of the London WRZ is to events of two consecutive dry summers with an intervening dry winter. SESRO is of particular benefit during events of this duration, and the proposed release volumes have been tuned to ensure that the scheme would have maximum benefit when considering the vulnerabilities of the existing London supply system.*

I agree that assessment of deployable output should use a 'system approach' and that some supply systems may be vulnerable to short droughts, for example direct river abstractions with no reservoir storage, whereas others may be vulnerable to very long droughts, particularly if the reservoir storage is large in relation to the water available for re-filling – this already applies to the existing London supply system and will apply even more if Abingdon reservoir is added to the supply system, with no additional source of re-fill.

Thames Water is incorrect to say that "*GARD's focus is solely on the Deployable Output of the reservoir".* GARD's modelling simulates the entire London and SWOX supply system in the same way as Thames Water's WARMS, Aquator and Pywr modelling. The deployable output changes assessed by GARD's modelling are for the whole supply system, not just for Abingdon reservoir.

Although I agree that the critical drought duration for the existing London supply system over the past 100 years has been two consecutive dry summers with an intervening dry winter, the critical drought duration is likely to change with the addition of Abingdon reservoir, thereby almost doubling the available system storage with no additional source of re-fill water. The threat of long duration droughts becomes greater if the system is expected to be resilient to 1 in 500 year events, rather than a 1 in 100 year or worst historic event.

Thames Water then goes on to reject GARD's use of WRSE's suggested approach of using artificial long drought events, which recognises that long droughts probably won't be generated by stochastic data trained on the 1950-97 period because it contained no severe long duration droughts:

From pages 115-116 of Appendix G2 of TW's SoR:

As a related point, the use of "dry" [as in creating artificial droughts by combining two dry but separate years] does not capture the range of "dryness" which could occur. If there is a 1 in 500-year drought event of 18 months duration, while longer droughts could occur, they would either be:

- *Equally dry (in terms of mm rainfall per month) for a longer duration, and thus more severe than a 1 in 500-year event (each month of very low rainfall being an unlikely event)*
- *Less dry (in terms of rainfall per month) and thus not as much of a risk for the existing London supplies*

GARD have presented modelling of an event which is a composite of 1996-97 preceding the 1933-34 event. According to the Standard Precipitation Index (12m accumulation period) data available from the UK CEH Water Resources Portal, the SPI-12 for the Thames catchment for Jul 96 to Jun 97 was minus 1.4, meaning that the event was a c.1 in 10-year event from a rainfall deficit perspective. Preceding the c.1 in 100-year

1933-34 event with a 1 in 10-year event means that this event could represent something more like a 1 in 1000-year event, i.e., an event less likely than we should consider in our planning.

As noted in our previous responses, we do not agree that the stochastic dataset under-represents long-duration droughts.

As also noted in a previous response, what GARD refer to as "advice" from WRSE is not advice, and is instead an allowance from WRSE to diverge from the preferred methodology should companies consider that a key vulnerability of their supply system is omitted. Our consideration is that the stochastic datasets properly consider drought events which may occur and to which our supply system is vulnerable.

Thames Water's rejection of GARD's example of the 1996-97 drought being followed by the 1933-34 drought is based on the argument that this would be a 1:1000 year occurrence

(1:10 for 1996-97 x 1:100 for 1933-34), and therefore more severe than the 1:500 year resilience standard. This argument misses the point that the 1996-97 drought being followed by the 1933-34 drought was simply an example of what could happen in artificially generated droughts based on changes in the sequencing of droughts of the past 100 years.

In GARD's Addendum and in Section 2.2.1 of this paper, I have provided evidence that the stochastic data generated different versions of the patterns of rainfall and drought sequences in the historic period 1950 to 1997, so was unable to generate the types of long drought, of more than 18 months duration, against which the London supplies have poor resilience if dependent on Abingdon reservoir. As I previously stated, this potential weakness in the stochastic data was identified in WRSE's Method Statement for Stochastic Climate Datasets[12]:

> *"As with any dataset generated based on existing datasets using statistical methods, the stochastic weather sequences are only as good as the datasets on which they are trained. As stated above, the stochastic dataset is formed of 400 48-year sequences and is trained on the 1950-1997 baseline period. There is a risk that extreme, extended droughts may not necessarily be well reflected in the dataset, although quantifying this risk is extremely difficult. Companies may complement the stochastic dataset with drought artificial weather series to represent prolonged drought events (which the stochastic generator will not have been trained on)."*

In my opinion, the example of the 1996-97 drought being followed by the 1933-34 drought shows that the London supply system, if dependent on Abingdon reservoir, would be vulnerable to droughts of three dry summers and two intervening dry winters. Therefore, there should be a proper investigation using artificial drought weather series as proposed in WRSE's method statement. It is not acceptable for Thames Water to refuse to do this on the grounds that "*Our consideration is that the stochastic datasets properly consider drought events which may occur and to which our supply system is vulnerable.*"

In view of the long running nature of the dispute over Abingdon reservoir resilience to long duration droughts and Thames Water's entrenched position, I think it is essential that this matter is now subject to a detailed and genuinely independent investigation before any decision is taken on major new supply sources.

[12] WRSE Method Statement on Stochastic Climate Datasets: Consultation Version, July 2020, paragraph 2.7

4. Deployable output and operational usage of the STT

4.1 Validity of Pywr STT modelling and deployable output of STT options

GARD's Addendum to the consultation on Thames Water's draft WRMP24 concluded that the Pywr modelling has grossly under-estimated the deployable output of unsupported STT options.

As commented upon in Section 2.3 of this paper, Appendix I of TW's WRMP24 provided validation details for Pywr modelling of existing London supplies but no validation of modelling of STT options. As our own check on Pywr model validity, in Section 4.1 of the Addendum we compared our modelling of the unsupported transfer using historic flows with Pywr modelling using stochastic flows with no climate change. This was shown in Table 8 on page 45 in the Addendum which is copied below:

	GARD DO for historic flow records	Thames Water DO for 1:100 return period
300 Ml/d USTT	129 Ml/d	90 Ml/d
500 Ml/d USTT	182 Ml/d	130 Ml/d

Note: Thames Water DOs are estimated from the plot in Figure 23

Table 8 - Comparison of GARD and TW assessments of USTT deployable output

Our modelling showed that the 1:100 year DO of the unsupported 300 Ml/d transfer was 129 Ml/d compared to Thames Water's Pywr figure of about 90 Ml/d. For the 500 Ml/d unsupported transfer, we estimated the 1:100 year DO to be 182 Ml/d compared with Thames Water's Pywr figure of about 130 Ml/d. These are large differences which could affect the decision on whether to proceed with an early development of the STT aqueduct.

In the Addendum, we pointed out that the deployable output differences shown in Table 8 are highly significant because the unsupported transfer would be a viable first phase of the STT, not dependent on the Minworth or Vyrnwy support sources, and the additional London deployable output would allow all the Chilterns chalk stream abstraction reductions to go ahead as soon as the Severn to Thames aqueduct is built, potentially in the early 2030s. It would also provide insurance against the planned PCC and leakage reductions not materialising over the next 10 years.

Thames Water's response to this criticism in Appendix G2 to the SoR is shown below:

From page 119 of Appendix G2 of TW's SoR:

GARD have undertaken modelling to identify a Deployable Output using historical flow data and not including the impacts of climate change. A "worst historical" Deployable Output assessment is not appropriate for ascertaining the DO benefit of the STT when planning for "1 in 500-year" conditions, and it is also not appropriate to ignore the impact of climate change in this case. As such, GARD's modelling is inadequate for the production of Deployable Output benefit figures for the STT for use in WRMP24. The modelling undertaken to produce the DO benefit values adopted in our WRMP is robust.

TW's response misses the point of the DO comparison in the Addendum Table 8 copied above and does not address GARD's criticism. The DO values in the Addendum Table 8 are GARD's assessment of DOs using historic flows since 1920 (ie no climate change) compared with Thames Water's assessment with Pywr modelling using stochastic flows without climate change. Assuming the worst historic drought since 1920 to be about a 1:100 year event, this is a like-for-like comparison which tests the validity of the Pywr modelling. It shows that the Pywr modelling appears to have substantially under-estimated DO gains arising from unsupported transfers.

To address GARD's criticism Thames Water should be required to provide full details of:

- Aquator model output of the unsupported transfer using historic flows derived from historic weather and river gauge data, as for WARMS2
- Equivalent Pywr model output of the unsupported transfer using historic flows derived using the Pywr hydrological modelling
- Explanation of a) any significant differences between the Pywr and Aquator modelling and b) any significant differences between the DOs assessed using historic flows and the 1:100 year DOs assessed using stochastic flows (ie the equivalent comparison to that shown in Table 8 of our Addendum).

4.2 The need for UU replacement sources for Vyrnwy support option

Thames Water appear to have assumed that at least 80% of the nominal support from Vyrnwy reservoir will require replacement of deployable output through new United Utilities sources, with high associated costs which make this option much less attractive. GARD's modelling shows that only about 50% replacement deployable output is needed. This would mean that the costs of STT options with Vyrnwy support may have been inflated by the cost of up to about 70 Ml/d of unnecessary replacement sources.

GARD's evidence in support of this is shown below in Extract 8 from the Addendum:

Extract 8 from GARD's Addendum pages 46-47

4.2 Need for UU replacement sources for Vyrnwy support option

We have found no clear statement in the WRMP documentation or Gate 2 reports for the amount of United Utilities replacement sources needed if Vyrnwy reservoir is used to support the STT. The Gate 2 feasibility report on the North West transfer refers to the replacement sources as 'sub-options' and recognises that they will not be needed continuously[44]:

> *"Allowing this indirect type of trading support helped us to reduce the capacity of [sub-] options required for trading well below the total transfer amount (167 Ml/d versus 205 Ml/d)."*

However, this is still suggesting that about 80% of the nominal Vyrnwy support amount will need to be replaced by deployable output from the new sub-options. GARD's modelling shows that this is not the case, as illustrated below in plots of modelled operation of Vyrnwy reservoir in the droughts of 1933/34 and 1975/76:

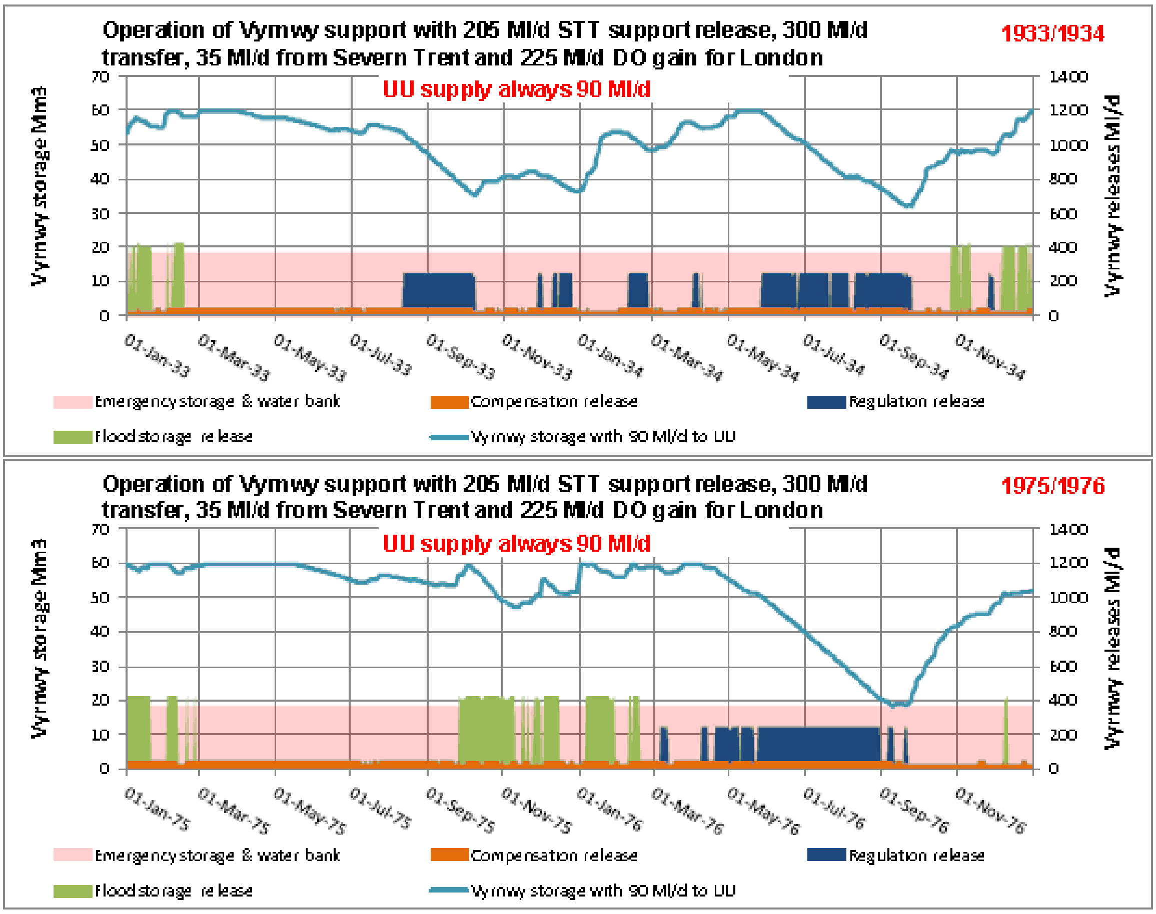


Figure 26 - Potential UU supplies from Vyrnwy while supporting STT

[44] NW Transfer Gate 2 feasibility report, paragraph 4.17

Extract 8 from GARD's Addendum pages 46-47 continued …

> The upper plot shows that in the drought of 1933/34, usually the critical historic drought for the reservoir when supplying United Utilities, it can easily sustain a continuous supply to United Utilities of 90 Ml/d, as well as providing 205 Ml/d of support to the STT when needed. The lower plot in Figure 26 shows that 90 Ml/d is the deployable output to UU that can just be sustained in the 1975/76 drought, in addition to the 205Ml/d support to the STT – in this scenario, 1975/76 is the critical historic drought.
>
> The deployable output of Vyrnwy reservoir when used for continuous direct supply to United Utilities is about 190 Ml/d (from GARD's modelling). Therefore, the sub-options needed to replace United Utilities deployable output lost when providing 205 Ml/d of support to the STT are only about 100 Ml/d (190 Ml/d less 90 Ml/d), not 167 Ml/d as stated in the NW Transfer feasibility report.
>
> In the WRMP documentation and the Gate 2 reports we have founded no statement of the amount of sub-option replacement sources assumed when costing STT options with Vyrnwy support. If options with 205 Ml/d of Vyrnwy support have assumed 167Ml/d of sub-options are needed, rather than 90 Ml/d as shown by our modelling, the STT option cost will have been inflated by the capital and operating costs of an unnecessary 77 Ml/d of replacement sources. If the costing assumption has been 1:1 replacement sources for the nominal amount of Vyrnwy support, as we suspect has been the case, the inflation of the STT option costs will have been even higher. The lack of clarity of the assumed amount of replacement sources is a major failure of transparency in the WRMP and Gate 2 documents.

Thames Water's response to this detailed and evidence-base criticism is shown below:

From page 119 of Appendix G2 of TW's SoR:

> *Thames Water have not made any assumptions about the resource which United Utilities would need to invest in, in order to replace water traded to facilitate a supported Severn Thames Transfer. These calculations have been undertaken by United Utilities. Instead, United Utilities have provided Thames with prices for water from Vyrnwy, which incorporate the need to invest in new sources, making use of utilisation series from WRSE (which have been provided to GARD). We have not, however, been provided with information regarding the derivation of these prices for reasons of commercial confidentiality.*
>
> *Our consideration is that this is an issue which should have been raised in relation to the United Utilities WRMP24 consultation.*

In other words, Thames Water have not questioned United Utilities' assessment of the need for replacement sources, despite the criticism raised by GARD in Extract 8. Instead GARD have been criticised for not raising the matter directly with United Utilities. In my opinion this is unacceptable and United Utilities and Thames Water together should be required to address GARD's criticism in a transparent manner, with proper supporting evidence, including modelling equivalent to that undertaken by GARD in Extract 8.

4.3 Operational usage of the STT

On page 47 of the Addendum, GARD said that In the WRMP and Gate 2 documents we had found no statement of the assumed annual amounts of STT operation for assessing operating costs. No time series data had been supplied for Pywr modelling of operational use of the STT. This is another failure of transparency, particularly as the high pumping costs, energy use and carbon impact of the STT are frequently touted as factors against the scheme.

From page 121 of Appendix G2 of TW's SoR:

Operational costs of the STT are presented in the WRMP Tables, including fixed opex (£/yr) and variable opex (£/Ml). These figures are used in our WRMP programme appraisal. Appendix W of the dWRMP included weightings applied to different scenarios considered in the investment modelling.

We note that elsewhere in GARD's response they have criticised the utilisation assumptions which they have found in the Gate 2 reports, while here they criticise not being able to find the same utilisation assumptions.

We note that GARD have been sent detailed pywr model data:

- *Flows for the Severn at Deerhurst for the stochastic river flow timeseries*
- *Control Curve Crossing data for DO calculations from an unsupported and fully supported STT*
- *A timeseries of utilisation from an STT DO run using the full stochastic dataset*

I now realise that some Pywr model output of STT usage had indeed been provided in the suite of data files in EIR-21-22-749 and I apologise for the oversight.

The modelling of the utilisation is described in the extract below from EIR-21-22-749 and the result is summarised in the extract copied from the Pywr output file "Utilisation_STT_300_200.xlsx" show usage of a 300 Ml/d transfer with 200 Ml/d of support sources:

From EIR-21-22-749:

This dataset is a timeseries of Severn-Thames Transfer utilisation from a WRSE Pywr model run investigating the utilisation of the Severn-Thames Transfer under 'DO-level demand' (1 in 500-year DO demand with the scheme in place) conditions.

You will note that there are timeseries of 'Total', 'Unsupported' and 'Support' transfers. In order to assess the utilisation of support (which is used in determining how much support can be made available and determining option cost), separate nodes have been included to capture transfers that would be possible without support being available, as well as transfers for which support would be needed.

In this model run, the scheme being investigated was a 300Ml/d pipeline, with 200Ml/d of support available at Deerhurst (i.e., this support could be from any source and would inclusive of losses in the Severn), and the demand placed on the model was approximately equal to the 1:500-year DO with the scheme in place.

Other notes about this run:

- Note that these outputs are from a run which was carried out several months ago. We are continuing to work on DO modelling as part of work to produce the WRSE Best Value Plan and SRO Gate 2 submissions.
- Demand savings were switched off in this run, as we were interested in utilisation under 'DO' conditions.
- Losses can be applied to the timeseries of transfer, but these were not captured. These would be 2% losses between Deerhurst and Culham, and 2% losses between Culham and Teddington. In addition, a 4-day travel time is assumed between Culham and Teddington.

From Pywr output file "Utilisation_STT_300_200.xlsx"

	Transfer	Unsupported Transfer	Support needed
Total Days	1,002,400	563,260	529,824
% of time	14.3%	8.0%	7.6%

The overall utilisation of 14% is similar to that shown by GARD's modelling using historic river flow data and seems reasonable. However, this usage is much less that stated in the Gate 2 report for the STT as shown in the extract below:

From Pywr output file "Utilisation_STT_300_200.xlsx"

4.17 The original source of this utilisation time series is based on DO modelling completed with the Thames Water Aquator model using historic river flow values. We advanced this assessment of utilisation by using results from stochastic DO modelling completed with the up to date WRSE Pywr model. Table 4-3 provides a summary of frequency, duration and magnitude of STT utilisation based on historic and stochastic DO modelling. On any day, the water conveyed by the STT Interconnector could be either all unsupported flow, only support options or a combination of the two. There is a similar pattern of overall utilisation of the unsupported transfer and the support options for the historically derived and stochastically derived utilisation patterns.

Table 4-3. Summary of utilisation over historic and stochastic time series of River Severn Flow

Aspect	Based on historical flow data (1920 – 2010)	Based on stochastically generated flow data (climate drivers from 1950 – 97)
Overall utilisation throughout the complete time series – unsupported transfer	6.20%	7.80%
Overall utilisation throughout the complete time series – all types of support	22.30%	22.60%
Period of support in key droughts Note: the realisation number represents one version of the stochastic sequence	Top 5 historical	1 in 500-year droughts (as highlighted by WRSE)
	244 days (1944)	230 days (realisation 66, 1976)
	234 days (1921-22)	232 days (realisation 152, 1976)
	226 days (1976)	194 days (realisation 209, 1992)
	214 days (1990-91)	209 days (realisation 302, 1976)
	197 days (1945)	189 days (realisation 348, 1992)

It appears that the utilisation assumed in the Gate 2 costings, and we presume in WRSE's option comparisons and WRMP conclusions, was based 22% rather than 14% as per the Pywr model output supplied with EIR-21-22-749.

I conclude that Thames Water should be required to provide full and transparent details of the utilisation assumptions inherent in evaluating STT options, including the relevant Pywr output. This information should be scrutinised by the regulators as well as being made available to GARD.

5. Ofwat's response to GARD's concerns about deployable output assessments

GARD's Addendum to the response to Thames Water's WRMP24 was also included as an Addendum to GARD's response to Ofwat's draft Gate 2 decisions to ensure that RAPID and the regulators were fully aware of GARD's major concerns over Thames Water's assessment of the deployable outputs of the Abingdon reservoir and STT options.

5.1 Recognition of GARD's concerns in final Gate 2 Decisions

Ofwat's final response to the Gate 2 decision on Abingdon reservoir (SESRO) recognised GARD's concerns as shown in the extracts below:

From pages 16-20 of Ofwat's final Gate 2 decision report for SESRO:

Deployable output and drought resilience

- GARD believe the stochastic river flow data and Pywr modelling are not fit for the purpose of assessing the deployable output and drought resilience of Abingdon reservoir (SESRO).
- Stochastic data excludes long droughts and Thames Wate haven't considered artificial weather series to account for prolonged droughts.
- Long duration droughts are likely to reduce Deployable Output of SESRO as shown by some of GARD's modelling.

- The proposed Abingdon reservoir still only allows 6% of emergency storage, as compared to typically 20% for other major UK reservoirs. The last 6% of water will probably be of very poor water quality and is likely to be unusable.

- GARD assert that there is no serious estimate of the time taken to fill the reservoir after completion of construction. What little detail there is contains a major error through using a Culham minimum required flow of 450Ml/d instead of 1450Ml/d. The absence of probabilistic estimates of times needed for initial filling of the reservoir is a major weakness in the gate two reporting.

Recommendations for gate three activities

- GARD think that Ofwat's recommendations for gate three actions should specifically cover the deficiencies in the gate two reports. These cover:

2. Provisions for dead and emergency storage and the acceptability of water quality in the reservoir at times of extreme drought and near-emergency drawdown.

7. Independent expert review of the stochastic data and Pywr modelling used to determine Abingdon reservoir and drought resilience.

In addition, Ofwat's final response to the Gate 2 decision on the STT recognised GARD's concerns about the deployable output of the STT as below:

From page 18 of Ofwat's final Gate 2 decision report for STT:

- GARD assert that STT deployable output is underestimated. They propose that Ofwat's gate two decision report should state that the independent review they have advocated for the stochastic data and Pywr modelling of Abingdon reservoir should include the assessment of deployable output of the unsupported STT. This should form part of the evidence needed for the interim STT checkpoint that they have advocated in Section 4.2.

The above extracts from Ofwat's final Gate decision reports, have acknowledged GARD's concerns about the validity of the deployable output assessments for Abingdon reservoir and the STT.

5.2 Ofwat's proposed actions to deal with GARD's concerns

In the final Gate 2 decision report for Abingdon reservoir (SESRO), Ofwat put forward the following response to GARD's concerns:

From pages 35-36 and 37 of Ofwat's final Gate 2 decision report for SESRO:

3.2.19 Deployable output and drought resilience

The use of stochastic flow data reflects the requirement to test droughts larger than those observed in the historic record, such as drought events with 1:500 year return periods. Solution generation of stochastic flow data is expected to follow Water Resource Planning Guidelines Supplementary Guidance: Planning to be resilient to a 1 in 500 drought (England), and Supplementary Guidance: Stochastics.

GARD's concerns around lack of assessment of long duration droughts and the impact on deployable output of Abingdon reservoir are also included in GARD's response to the Water Resources South East consultation. As these matters concern water resources planning, it is the responsibility of WRSE and Thames Water to answer these queries as part of the consultation response process.

3.2.24 Independent expert review

The areas raised by GARD are within the ambit of the town and country planning process or other statutory controls. To satisfy the requirements of these controls, the solution owners will need to satisfy the independent regulatory authorities which governs these processes. This independent regulation will address the concerns raised by GARD.

Appendix A to Ofwat's final Gate 2 report contains no required actions for the water companies to undertake to address the concerns raised by GARD.

In the final Gate 2 decision report for the STT, Ofwat put forward the following response to GARD's concerns:

From page 22 of Ofwat's final Gate 2 decision report for the STT:

3.2.7 Solution design

There were concerns from one stakeholder around stochastic data. We expect all technical work and modelling to have undergone review and quality assurance. Activities should follow best practice guidance where relevant, and to state this in submissions. Specifically on stochastic data, consultant investigations have been commissioned by the regional groups that have reported on comparisons of stochastic data sets, including those used by the regional groups, and alternatives. The regional groups have also held workshops for wider stakeholders on methods used and have made data available for wider stakeholder use through these workshops.

Appendix A to Ofwat's final decision report for the STT contains no actions to address GARD's concerns about the deployable output and operational utilisation of the STT.

In my opinion, Ofwat have not adequately dealt with the concerns raised by GARD. The only actions proposed are (from the excerpts on the previous page):

1. *As these matters concern water resources planning, it is the responsibility of WRSE and Thames Water to answer these queries as part of the consultation response process*.
2. *The areas raised by GARD are within the ambit of the town and country planning process or other statutory controls. To satisfy the requirements of these controls, the solution owners will need to satisfy the independent regulatory authorities which governs these processes. This independent regulation will address the concerns raised by GARD.*

I appreciate that Ofwat's final Gate 2 decision reports were published in June 2023 before publication of Thames Water's statement of response at the end of September 2023. However, it is now apparent from the analysis of Thames Water's statement of response presented in this paper that they are unwilling or unable to make proper evidence-based responses to GARD's concerns. In my opinion, the views of Thames Water and their consultants on these matters are now so entrenched that they will never concede the need to make any changes, regardless of the evidence presented.

Therefore, I think the time has come for the independent expert view called for by GARD. It is not clear whether the statement above "*This independent regulation will address the concerns raised by GARD"* refers to regulation of planning consent for major schemes or to regulation of the Water Resource Management Plans by Ofwat and the Environment Agency. However, in my opinion the independent review needs to be undertaken before any decisions are made on acceptance of the final WRMPs.

John Lawson, FREng, FICE, FCIWEM
6th October 2023