Reviewer's comments in *italics*; authors' response in plain text.

This is an interesting manuscript that contributes to a growing body of research that applies probabilistic climate information and resource system analysis to determine efficacy of adaptation options. The Thames Basin, UK has been the subject of numerous climate risk assessments. However, the paper adds new insights on the significance of future water demand relative to other key uncertainties. The authors have also applied end-user relevant water resource indicators, and introduced technical advances in the form of a new spatial weather generator. Some methodological details require further elaboration (e.g. treatment of change factors, groundwater exploitation, return flows) and the assumptions behind the water demand estimates could be more transparent.

The plausibility of some adaptation options (e.g. 35% demand reduction by 2020) is highly questionable and, arguably the most serious threat to the system – multi-season droughts – warrants further consideration.

On this basis, publication is recommended subject to the following modest revisions, minor corrections and clarifications.

We thank R Wilby for his overall positive comments on this manuscript and for the valuable comments and challenges that would strengthen and improve this paper. We respond to the comments below, indicating what changes or additions we would make if invited to prepare a final manuscript.

Main comments

[Abstract] Please incorporate a few more headline statistics and link any projected impacts to the specified time horizon. Presumably, the cited five-fold increase in drought order occurrences pertain to the 2050s?

Yes, the five-fold increase mentioned in the abstract refers to results from the 2050s. We could also include some results for the 2020s for example a decrease in per capita demand of 3.75% reduces the median frequency of DS4 measures by 50%.

[Methodology/Results] The UKCP09 projections suggest an average 18% reduction in summer precipitation and 15% increase in winter for the Thames Basin. However, a fundamental limitation of the change factor methodology is that it does not readily produce hitherto unseen sequences of exceptionally dry periods that would truly stress the water supply system (such as two or more dry winters and summers in a row). It would be helpful to know what the longest below average sequence was generated by the method of adjusting observed rainfall and refitting the STNSRP to the perturbed series.

We agree with the reviewer in that sequences of dry periods are important and potentially critical to water supply systems. Mean seasonal precipitation was calculated for spring, summer, autumn and winter and then each season was expressed as an anomaly from the means. The figures for the longest sequences of negative seasonal anomalies are 10 for the control and 18 for SCN50, a significant increase.

[Methodology] Given the heterogeneity of land cover in the Thames basin, the decision to apply a single Potential Evapotranspiration (PET) record at the centroid of the basin should be further justified. For instance, how representative is this of the moisture loss from the urban sub-catchment?

We recognise that the calculation of PET is essential in calculating future water resource availability and that different downscaling methods yield different PET change predictions. The Thames basin was divided into three sub-catchments to conduct the hydrological modelling as shown in Figure 2.

Within each of the catchments there are a variety of land uses which would in turn affect moisture losses, as the reviewer highlights. Given the similarity in the three catchments in terms of elevation and heterogeneity of land cover, and that CATCHMOD is a lumped model, only one PET series for each sub-catchment could be used as input, it was decided to use the same PET record for the three areas. Unfortunately with applying a lumped model it is not possible to investigate how representative this is for the urban sub-catchment. However, in a publication forthcoming we use a physically-based spatially distributed hydrological model which uses spatially correlated precipitation and PET input data from the UKCP09-WG on a 5km grid. Thereby giving a better representation of input data across the Thames basin.

[Methodology] Explain how groundwater abstraction was handled by the catchment/ water resource modelling. What input data were used for the groundwater source node shown in Figure 4? Likewise, what values were applied for return flows to the Thames from the water treatment works node?

Groundwater is included as an aggregate inflow of 467.4 Ml/d directly into meet London's demand. A proportion of the inflow to the water treatment works is leakage, this is equivalent to 12% of demand and is returned back to the river and modelled as a contribution to the minimum environmental flow. This information could be included in the caption for Figure 4.

[Methodology/Results] How well does the LARaWaRM replicate the historic frequency of demand saving days when given observed climate and water use profiles?

As far as we are aware there is no data indicating the historical frequency of demand saving days for the catchment. Furthermore there is also no account of the vast changes that have taken place in the catchment around usage data, given the shift from industrial use to predominantly residential or of the supply infrastructure changes over time. We have run the observed flow record (1883-2005) through LARaWaRM which gives a single value that falls within the range of those produced using control climate scenario flows. This value could be included in the caption of Figure 8.

[Methodology] Provide a summary table of all the demand measures that were factored into the analysis whether implicitly or explicitly. Explain how demand per capita was altered to reflect technological advances and efficiency measures. State the assumed consumptive water demand (i.e. fraction of used water that is not returned to the Thames as treated effluent).

We agree a summary table of the values and ranges mentioned throughout the text will be will be useful and would add as a table in Supplementary Material.

[Results] The 10th percentile change in precipitation shown in Figure 5 is based on the ensemble uncertainty. However, what proportion of weather generator runs produced a decrease in precipitation in all months? This is no doubt a much rarer likelihood than 10% of ensemble members. In other words, 10th percentile scenario (and attendant adaptations) is actually much rarer than implied for a continuous simulation.

We calculated that the monthly means for the control and 2050s to determine whether the monthly mean increased or decreased relative to the control. The data showed that none of the 100 runs show a decrease in every month - so the likelihood is much rarer than 10% as the reviewer suggests.

[Discussion] As noted before, further comment is needed on the treatment and effectiveness of the adaptation measures under conditions of multi-season drought. For instance, how more/less frequent are 24-month periods with below average precipitation under SCN20 and SCN50? Would the preferred mix of adaptation measures differ if these particular climate threats are considered?

We have calculated the number of spells of below average rainfall that last at least 24 months (i.e. at least 8 seasons of negative anomalies) for one grid cell. In the control period this occurs you typically get one such spell (ranging from 1-2 occurrences). In SCN50, we start to see evidence of more frequent occurrences up to 5 (ranging from 1-5 occurrences). In the UK groundwater and reservoir recharge typically occurs over the winter from November to April. Successive dry winters cause significant water resource issues particularly in the east and south of England due to the importance of groundwater in providing base flows in rivers. As the reviewer suggests this may have implications for the mix of adaptations, with a greater emphasis needing to be applied to for example, demand-saving measures, leakage reduction and desalination. Considering multi-season drought alongside sequencing of adaptation options as proposed by e.g. Beh et al. 2015 or adaptation pathways e.g. Haasnoot et al. 2014 would be a valuable extension of this work. This would be reflected in the conclusions.

[Conclusion] Please add a paragraph spelling out the priorities for further research.

The following paragraph would be added to the end of the paper highlighting priorities for further research.

This study of the Thames catchment and subsequent analysis has highlighted the following priorities for future research. Firstly, which will be addressed in a forthcoming paper is an extension of the climate scenarios to include the 2080s time period, coupled with the application of a the spatial weather generator feeding a physically based, spatially distributed hydrological model which will allow better representation of both the climatological and land cover heterogeneity of the catchment. Furthermore, it will enable changes in land cover i.e. increasing urban areas to be considered. Secondly, recognising the importance of groundwater in the Thames catchment and hence the potential impact that multi-season droughts may have on the area, further research is needed to understand how such trends may affect or influence choice of adaptation options. This alongside a third research priority looking more generally about the sequencing of choice of adaptation options over indicative planning horizons taking account of trade-offs with reducing greenhouse gas emissions or investment portfolios could make use of more robust decision making frameworks under uncertainty such as those proposed by for example Beh et al. 2015 or Haasnoot et al. 2014.

Minor corrections and clarifications

[P8856, L10] Exactly what "changes" in upland river catchments were analysed?

Changes in precipitation and PET; this would be made clear.

[P8856, L17] ": : :greater proportion of uncertainty: : :" by when? Note that the relative importance of different uncertainty components varies with time in the future i.e. climate variability (near-term), climate model (medium-term), emissions scenario (longterm). Hence, it is critical to attach the time-stamp to any such impact or uncertainty statement.

The work referred to here was considering scenarios in the 2080s – the time stamp would be added, taking note of the reviewer's comment about the importance of uncertainty components varying over time.

[P8857, L25] In what ways are the PET data "consistent"?

The phrase 'consistent' refers to the relationship with the rainfall scenario. 'Corresponding' is perhaps a better term.

[P8858, L1] Ditto for "consistent river flows"?

Here the term 'consistent' refers to the relationship of flows to the rainfall and PET series. Again 'corresponding' may be a better term.

[P8858, L20 and P8871, L26] Given the large uncertainties, better to use "could" rather than "will".

This change would be made.

[P8859, L19 and P8870, L29] Comment: Given that the longest time horizon investigated was the 2050s, uncertainty due to emissions scenario can be largely discounted.

[P8860, L7] At what time-scale were the change factors produced? Was it annual, seasonal, or monthly? Please specify.

Change factors were monthly. This would be clarified.

[P8861, L5] Please cite the Nash-Sutcliffe efficiency (and other performance metrics) for the validation period.

These would be added.

[P8865, L23] Justify the assumption that per capita demand remains the same as present in SCN20 and SCN50.

In the scenarios presented in Figure 8 we are concerned with determining the individual impact of climate projections and population projections, therefore per capita demand is kept constant. The scenarios presented in Figure 9 go on to consider per capita demand.

[P8865, L25] The existing level of service for severe water rationing (DS4) is set at zero frequency (Table 1). What criteria were used to determine when the DS4 measure is invoked under SCN20 and SCN50?

The criteria remains the same as the present day. As suggested in the reviewer's comment on Table 1 below, the trigger thresholds will be added as a column to Table 1.

[P8866, L22-24 and P8872] Please explain how a 100% reduction in the mean number of DS4 days can be achieved twice over (first by leakage reduction and a new reservoir, second by adding a desalination plant to the portfolio).

This figure has been incorrectly reported. It should read 37.5% and would be changed.

[P8868, L22] Should Fig.10c be Fig.8c?

Yes. This would be amended.

[P8868, L29] Comment: Water use for industry and mining may have declined, but fracking of natural gas reserves in southeast England could add to future water demand.

[P8869, L10] How sustainable is the expected 10-15% water saving in metered households? Some studies suggest that demand creeps up in the long run.

It is difficult to comment on how sustainable the expected 10-15% reduction in water usage would be. However, this could be monitored and analysed from meter data over time. Technological advances in water appliances may indeed offset any increases.

[P8869, L18] Note that a 35% reduction in per capita demand by 2020 is unrealistic. It is much better to reframe these results as a sensitivity analysis that demonstrates the scale of the challenge ahead.

We agree and on P8862, L28 we refer to this part of the analysis as a sensitivity test and on P8869, L18 highlight that such a reduction is unrealistic.

[P8869-P8870] Note that the water supply-demand outlook could be even more finely balanced if more stringent treatment of environmental flow requirements are taken into account. This factor might be particularly problematic under the multi-season drought episodes noted above.

[P8871, L23] Note that climate change impacts on water balance risks might exceed those due to population growth beyond 2050.

[P8872, L19] Please clarify the sentence "Given the typical time: : :"

This refers to the investment time scale of water resource planning; this would be clarified.

[Table 1] Add a column of the precise triggers/thresholds used by Thames Water to invoke the different levels of restriction.

The reviewer suggests adding a column to indicate the thresholds for demand saving measure. Given these thresholds are dynamic and vary monthly, and that it would not be helpful to include an average value we suggest that the table below is included as Supplementary Material.

Table X: Percentage of total storage capacity that invokes the different levels of restrictions

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Level 1	69.5	82.0	89.5	92.1	92.1	90.9	86.0	74.9	64.8	61.0	59.8	61.0
Level 2	42.2	47.8	55.3	65.2	74.0	76.5	74.7	67.7	55.7	46.2	41.1	39.8
Level 3	33.6	37.3	43.4	52.1	60.9	65.9	67.2	62.4	51.2	41.2	34.9	32.4
Level 4	17.4	19.8	24.8	29.8	34.8	38.6	39.8	39.8	37.4	31.2	25.0	20.0

[Figure 1] Please add boxes/arrows for "Observed data" and "Performance metrics" to complete the picture of the whole analytical framework.

Thank you for this suggest; these two boxes and relevant arrows will be added to Figure 1.

[Figure 3] Add the Nash-Sutcliffe scores to each panel.

These will be added.
Teddington Weir: Calibration NS 0.88. Validation NS 0.86.
Days Weir: Calibration NS 0.86. Validation NS 0.90.
Feildes Weir: Calibration NS 0.68. Validation NS 0.69.

[Figure 10] For clarity, replace "No" with "No supply measures", and spell out "D", etc. on the x-axis.

The abbreviations the x-axis on Figure 10 will be presented in full if possible given font size requirements. The abbreviations are explained in the figure caption.