Marine report

Id	Section	Comment	Reviewer initials	MO response	Post Moderation Meeting Met Office comments
6	2.1 The Sea Level “Jigsaw Puzzle”	Gravity effect could be included in this figure	Prof Bart Van Den Hurk	The effects of gravitational changes related to changes in ice and/or water mass re-distribution in the Earth System are discussed in lines 11-17 of page 9. Since there are also spatial patterns associated with changes in ocean circulation / density and GIA, we prefer not to highlight the gravity effect in particular (which is only one of the 3 components related to mass changes, the others being the local lithosphere response and changes in Earth’s rotation).	
10	3.1 Projections of Time-Mean Sea Level Change	A short explanation of the difference between North and South UK (role of gravity in Greenland, rol of GIA) could have been added to summary	Prof Bart Van Den Hurk	We feel that this level of detail is too much for a section summary, which is designed to provide a concise “top level” overview of the key results. The details are discussed in the main text.	
13	3.1 Projections of Time-Mean Sea Level Change	referring to "baseline period": how do you make the match to observed SLR?	Prof Bart Van Den Hurk	We have included an additional figure in this section to show the comparison between the UKCP18 GMSL sea level projections and the satellite altimeter observations.	
15	3.1 Projections of Time-Mean Sea Level Change	General: what kind of research would be needed to substantiate the findings and uncertainties regarding the Antarctica stability?	Prof Bart Van Den Hurk	We have added the following paragraph in section 3.1.2. However, we would be very keen to get input from the Peer Review Panel on the proposed text: “The potential for accelerated dynamic ice discharge over the coming decades and centuries emphasises the need for close monitoring and improved modelling of the West Antarctic Ice Sheet. In particular, efforts to estimate the ice sheet mass balance (Shepherd et al, 2018) must be continued, ideally supplemented with observational campaigns to monitor the ocean conditions under the ice shelves. Marine ice cliff instability has been proposed as an important potential feedback (DeConto and Pollard, 2016) and further research is required to strengthen the observational evidence for, and prevalence of, this mechanism. Ice models must continue to be improved with an emphasis on process-level understanding and making use of observational constraints when developing new projections.”	The PRP had no additional comments on this proposed text at the Moderation Meeting.
24	3.2 Projections of Change in Storm Surge and Extreme Water Levels	Did you also analyse the thousands of years of available seasonal forecast archives to get more grip on the uncertainty range?	Prof Bart Van Den Hurk	We acknowledge the value of forecast archives in constraining the uncertainty and we may be able to look at this approach in forthcoming work.	
26	3.2 Projections of Change in Storm Surge and Extreme Water Levels	The results in this figure are difficult to understand. What can be implied from this negative correlation? And how do the other 3 models explored fit in this picture?	Prof Bart Van Den Hurk	Please see also response to comment Id 92. The main point of this section is to illustrate why we chose to assess trends over a shorter period than we did in UKCP09. This section does not have any action for the user, so we have moved this section to appendix 1. Re. negative correlation: see below. Re. your question about the other 3 models: they do not have any clear signal (just a noisy 'cloud' around the origin).	Following further feedback at the moderation meeting, this section has been removed and replaced with a very brief text description of why we used a shorter period than we did in UKCP09.
29	3.2 Projections of Change in Storm Surge and Extreme Water Levels	Is the difference in the value of CRF in GC3.1 quantitatively consistent with this being the/a major factor in the difference in its ECS?	Prof Bart Van Den Hurk	We have added a panel to figure 3.2.6 showing the present day return level uncertainty for the four example locations. However, combining this source of uncertainty with the uncertainty in the projections of mean sea level change in a meaningful way is not a straightforward task, and further scientific work is required to establish the best way to do this.	Figure 3.2.6 now renamed figure 3.2.5
37	3.3 Projections of Changes in Wave Climate	Would be good to see the GCM results on a somewhat larger domain (e.g. entire Northern Hemisphere)	Prof Bart Van Den Hurk	We prefer to keep the emphasis on the UK and some of the details are lost when presenting a larger domain. The wider European domain is discussed in the associated publication by Bricheno & Wolf, which is cited in this section.	
39	3.3 Projections of Changes in Wave Climate	This southward shift is opposite to my general understanding of an expansion of the Hadley call associated with a northward shift	Prof Bart Van Den Hurk	We have added text to note that the southward movement of the jet is at odds with the expectation for a poleward movement of the mid-latitude jet, citing Barnes and Polvani (2013).	
41	3.3 Projections of Changes in Wave Climate	So SWH signal recovers in 2nd half of 21st century? Why would that be?	Prof Bart Van Den Hurk	The sentence now reads as follows: "Stronger reductions in the mean SWH are observed in 2030-2059 than in 2070-2099, which suggests that both low-frequency internal variability and climate change may be playing a role in shaping the simulated wave response."	
42	3.3 Projections of Changes in Wave Climate	How do the projections compare with the historic observations?	Prof Bart Van Den Hurk	An evaluation of the wave model performance with the available historical observations is presented in section A2.4. The wave observations are typically site specific and of limited duration. Hence the "hindcast" set from the wave model is the most useful way of establishing the potential changes in wave climate from the model projections.	
44	4.2 Projections of Time-Mean Sea Level Change to 2301	So in this scenario we don't see enhanced Antarctica mass loss?	Prof Bart Van Den Hurk	We have replaced this sentence with the following text: “While the central estimate shows a lesser contribution from Antarctica, owing to greater snow accumulated associated with a warmer atmosphere (and increased moisture transports), the large uncertainty in GMSL rise under RCP8.5 is dominated by uncertainty in the future dynamic ice discharge from Antarctica.”	

49	5. Comparison with UKCP09	Have these tabulated nrs been shown somewhere already? Would be good to have a place where all quantitative key results are shown together	Prof Bart Van Den Hurk	Our suggestion to add CMIP5 models to Strand 2 was motivated by a general expectation that including structural model diversity would broaden the range of realisations. This motivation related to regional *patterns* of change (for example, without CMIP5 models, Strand 2 would not include any realisations in which UK summer rainfall does not dry in the future), as well as global temperature outcomes. At the time of decision, we did have information from idealised experiments suggesting the possibility of some high-end global warming outcomes, but we didn't yet have the transient GC3.05-PPE simulations. Therefore, we don't think it would be appropriate to discuss the high warming earlier, in the *experimental design* part of section 3, as if it was the sole (or dominant) reason for adding the CMIP5-13 models. This is why we introduced the discussion of high warming in section 3.5, rather than earlier in section 3. We do, however, agree that the high warming result should be mentioned earlier in the report. We will address this by adding an Executive Summary at the front of the document, as suggested in comment #230. In this report, we are happy to include open and confident discussion that we can support with peer-reviewed evidence. However, we are constrained by the newness of the GC3.05 simulations and the project delivery timescale. We think it would be unwise to include "work in progress" information, due to the risk that some of this might later have to be retracted. In particular, we do not agree that the UKCP18 land science report is the right document in which to provide an extensive discussion of ECS. This is motivated by three reasons and follows discussion with the Met Office Chief Scientist on the issue of presenting ECS: • Work on understanding the ECS has not used GC3.05. Rather, it has used GC3.1 and UKESM1. • Aside from the standard model variant, GC3.05-PPE consists of members that each have different ECS values, but these values cannot be estimated from the scenario experiments available. • Work on understanding ECS in GC3.1 is not yet peer reviewed or published and is still evolving. Presenting a non-peer reviewed version risks being out of date and incomplete. The work will be published in due course and is relevant to multiple Met Office projects, including our CMIP6 contribution. However, we do agree with the need to address the high warming issue in the UKCP report and show how the PPE compares to CMIP5 models. We propose to do this by showing a comparison of the RCP8.5 warming response in the GC3.05 PPE with the IPCC AR5 simulations. This is a more appropriate comparison because it uses the scenarios actually applied in UKCP18. This marks a change from the draft report, in which the ECS estimate of 6.2K (for GC3.1) was included. We based	We added the following text, to give a sense of where ECS values for GC3.03-PPE may lie with respect to the current IPCC assessment: "In AR5, IPCC assessed ECS to have a likely range of 1.5-4.5ºC (Collins et al., 2013), and also judged thatthere is a small probability (of up to 10%) that ECS exceeds 6ºC. The levels of 21st century warming simulated for the RCP8.5 scenario suggest that most of the GC3.05-PPE members are likely to possess ECS values above 4.5ºC. ". At the Moderation Meeting, the PRP advised us not to add an Executive Summary, so we have not done this.
51	5.2 Projections of Extreme Coastal Still Water Levels	To be honest, I think there are too many degrees of experimental freedom shown in this section/plot	Prof Bart Van Den Hurk	We have simplified this section, showing only the results of UKCP09 and UKCP18, and summarising the causes of the difference. [TH]	
57	3.1 Projections of Time-Mean Sea Level Change	Section 3.1.1. and A1.1.1: Ice dynamics and Levermann (2014). As detailed in L14 the linear response approach is only valid for time-scales (much) shorter than the response time of the Ice sheet; it is not suited for self-sustained amplifications like Marine Ice Sheet Instability and it is tuned to coarse-resolution ice models that react slower and smoother than higher-resolution models and forced by coarse-resolution ocean models that lack the abrupt changes in under ice cavities simulated by higher-resolution models. Dot to these limitations the first author thinks the method is inadequate to make full projections (it is therefore not used in SROCC), as he sees it as a first step towards comprising the forcing uncertainty in sea-level estimates. The limitations should be more fully discussed. In addition the authors seem to favour the no-delay numbers, if these limitations are p;oestimates for deConto&Pollard 2016 in the emulated simulations of Edwards et al. (2018) (co-author), or using Golledge et al. 2015. Especially the reason why not taking the Golledge numbers into account should be motivated, as these contain explicit simulations of MISI where it occurs instead of a statistical approach that is very limited (L14).	Prof Sybren Drijfhout	As stated in section A1.1.1, Page 104, lines 10-11, the statistical fits to the Levermann et al (2014) results are based on their Table 6, i.e. the preferred choice of shelf models WITH time delay. In response to the reviewer's comment, we have included an additional paragraph in section A1.1.1 (see text below) and an additional figure, which presents a comparison between UKCP18, Golledge et al (2015) and DeConto & Pollard (2016). "One of the main limitations discussed by Levermann et al. (2014) is that the coarse resolution models used are not capable of representing the self-accelerating grounding line retreat associated with marine ice sheet instability (see section 3.1.2). Golledge et al (2015) have conducted ice sheet model simulations under RCP scenarios, but the computational expense associated with the higher resolution means that only two model simulations are available for each scenario (Figure A1.1.2). Therefore, Levermann et al (2014) was considered the best option available at the time the UKCP18 sea level projections were being developed. The UKCP18 central estimates do not capture the tendency for acceleration of ice mass loss seen in the Golledge et al (2015) simulations (Figure A1.1.2). However, the range of projections at 2100 presented by Golledge et al. falls well within the UKCP18 uncertainty estimates. The study of DeConto and Pollard (2016) shows similar values to UKCP18 under RCP2.6, but substantially larger values under RCP4.5 and RCP8.5 (Figure A1.1.2). The comparison with DeConto and Pollard (2016) is presented to provide further context for the UKCP18 in terms of the wider literature and the discussion presented in section 3.1.2."	The additional figure proposed here was further revised following useful discussions at the Moderation Meeting. The text in this section was further revised to make clearer the justification for using the Levermann et al (2014) results, in particular: (i) the more comprehensive (than Golledge) treatment of the uncertainties; (ii) the suitability of Levermann et al (2014) with the UKCP18 Monte Carlo approach to combining uncertainties from different sea level components.
58	3.1 Projections of Time-Mean Sea Level Change	Section 3.1.1 and A1.1.1: Uncertainty bands. The report mentions it follows AR5 in its method to calculate these. This is inconsistent with the use of a scenario-dependent estimate of ice dynamics and the statement on page 16; line 26-29. The ice dynamics estimate should be linearly (or partly linearly if a correlation value between 0 and 1 can be justified) added to other temperature-dependent processes like thermal expansion and SMB.	Prof Sybren Drijfhout	We are NOT following AR5 in respect of AR5 Antarctic Dynamics. At that stage there was no literature basis for quantifying/establishing/estimating the scenario dependence. This was discussed on Page 104 lines 3-19 of the Marine Report draft sent out for review. We have included the following additional paragraph in section A1.1.1 to discuss the issue of correlation between GMST and dynamic ice discharge from Antarctica: "An outcome of the parameterisation of the Levermann et al (2014) results is that the UKCP18 sea level projections do not represent any correlation between global surface temperature change dynamic ice sheet discharge from Antarctica. However, as discussed by Levermann et al (2014), the uncertainty in their projections was dominated by differences in the spatial pattern of warming across CMIP5 models (in particular how much warming was manifested in the Antarctic region), rather than differences across the RCP scenarios. A positive correlation between global surface temperature change and Antarctic dynamic ice discharge will tend to reduce the net contribution from Antarctica, through larger snow accumulation associated with a warmer atmosphere. In that sense, the lack of correlation in the UKCP18 projections will tend to promote more conservative (i.e. slightly larger) projections of the net contribution from Antarctica."	

59	A1.1 21st Century Regional Time-Mean Sea Level Projections	Section A1.1.2: Regression estimates for local sea-level response to global thermal expansion. Likely, CMIP5 models do not resolve and adequately describe the response on the shelf seas. At least this should have been discussed. Isn't there potential for dynamical downscaling with a regional ocean (shelf-sea) model?	Prof Sybren Drijfhout	This point is discussed in the following text on pages 107 and 108: "Following previous studies (e.g. Perrette et al, 2013; Bilbao et al, 2015), we estimate the oceanographic component of regional sea level by establishing linear regression relationships between local sea level and global thermal expansion in a number of CMIP5 climate models (Figure A1.1.2). These relationships vary both by CMIP5 climate model and geographic location around the UK, due to a spatial pattern of change that is highly uncertain (e.g. Pardaens et al, 2010; Slangen et al 2014). Since we cannot be confident in the ability of coarse resolution CMIP5 models to reliably estimate the spatial pattern of change in a single point within UK waters, we compute regression relationships for all UK coastal grid boxes for each CMIP5 model. During the Monte Carlo step, a CMIP5 model and coastal grid box is drawn at random to determine the local oceanographic sea level change by combining the regression slope with the time series of global thermal expansion." We are actively looking into the potential for dynamical downscaling with a regional shelf seas model. This work is being done in collaboration with Aimee Slangen and Tim Hermans (PhD student) at NIOZ in the Netherlands. At the time of devising the UKCP18 projections there were only two CMIP5-based simulations available, which was inadequate to characterise the model uncertainty. We do not include this discussion in the methods section since it seems superfluous / distracting to discuss an approach that wasn't taken. However, we will point out the potential for this approach in the Future Work section.	
61	3.2 Projections of Change in Storm Surge and Extreme Water Levels	Section 3.2 and 3.3: Changes in storm-surge and wave climate. I infer that these estimates are done without accounting for the possible effect of higher sea-level on storm surges and waves. This is OK when this effect is assumed to be small, but should be made explicit and motivated. I guess close to the coast in shallow water changes in water-depth can become O(1) and affect surge and especially wave climate?	Prof Sybren Drijfhout	This is a good point, and we should make it clear that an increased water depth will change the behaviour of surge and waves differently. With large wave events likely to be more damaging, when combined with a deeper total water depth (as there is less energy dissipation at the bed).	The following has been added to the list of caveats: "The primary effect of mean sea level increase on waves is to increase the mean height around which the waves fluctuate, leading to increased over-topping and coastal flooding. An important secondary inshore effect arises as follows. The maximum amplitude of waves before breaking in shallow water is limited by the water depth (e.g. Goda, 2000). Thus, an increase in mean sea level will in general have the secondary effect of moving the surf zone further inshore, increasing the wave energy available at the coast for over-topping and coastal erosion, thereby exacerbating the primary effect. We do not assess this secondary inshore effect here: our assessment of changes in the wave climate focusses on offshore wave changes. "
75	2. Introduction	What about atmospheric circulation changes?	Prof Mat Collins	Changes in the atmosphere drive ocean circulation changes - i.e. the sea surface height in balance with the ocean dynamics. To avoid potential confusion, we prefer not to introduce atmospheric circulation changes here explicitly.	
90	3.2 Projections of Change in Storm Surge and Extreme Water Levels	think you need to find a word other than "illustrative". To me that implies "probable", and section 3.2.2 has just shown any change in surge to be unlikely. I understand that we are now talking "high end" but it is still something of a contrivance to imlply it is illustrative. I also notice that throughout the report there is what statisticians would consider over-empahsis on the GFDL-ESM2M simulation; the truly parsimonious would treat it as an outlier - not keep coming back to it. The more believable mu = 0 gets one line on p28; the less credible high-end numbers get a whole page to themselves (p33). I think this presentation is misleading.	Prof Kevin Horsburgh	There is a precedent for 'illustrative', which was used in exactly this context in our Ice2Sea publication and our Singapore publication. Within this section, the word 'illustrative' is only ever used in combination with 'high end': 'illustrative high end'. Happy to be guided by the peer review panel consensus on this. Re. column-inches dedicated to GFDL-ESM2M: can see what you mean, but on the other hand, "no change" result is stated clearly up front in section summary on page 24, and page 25, 26, 27 and 28 are basically describing the "no change" finding. Also (of course) it doesn't take a whole page to say "zero at every port", whereas tabulating the high-end values at each port does. The central estimate of 'no change' is given clear prominence in the key findings.	
92	3.2 Projections of Change in Storm Surge and Extreme Water Levels	This analysis, sumarised by Figure 3.2.5, is unconvincing. That the modelled skew surge trend should reverse sign (putting completely to one side the fact that there is no observational evidence for either) sguggests strongly that it is not sampling the extremes for sufficiently long in th 1970-2005 period. This is all to do with stomriness of course. If the atmosphere suddenly behaves differently when the model swithces from hindcast to forecast then there is something fundamentally wrong with the configuration. The vague and unexplained nature of the scaling is a further concern.	Prof Kevin Horsburgh	Please see also response to comments Id 26 and 27. Regarding the scaling: The main point is that scaling does not change the important message of the plot. You could plot it without the scaling and it would look the same (apart from a change in the scale of the Y axis). For what it's worth, the scaling is based on (from memory): Standard deviation of slope behaves like period^(-3/2) under the null hypothesis. I'd have to find my notes at home for the full argument. I agree that it may not be sampling for long enough in 1970-2005, but if that was the end of the story, why the agreement between every port, in these two models? As you say, it presumably comes from storminess (large scale storminess, since different storm paths affect different ports, c/f Ivan's results). But regardless of whether the trends are due to large-scale 'noise' or global warming, I think this shows that we need to look at the 21st century extremes without the historical data included in order to identify potential 21st century trends. If it's just large-scale noise then no problem (though we still need to know how big the trends might be due to that noise). If, on the other hand, the model atmosphere behaves differently when the model switches from historical to future forcing (doubtful?) then that is an issue that is beyond the scope of this analysis.	Following further feedback at the moderation meeting, this section has been removed and replaced with a very brief text description of why we used a shorter period than we did in UKCP09.
94	3.2 Projections of Change in Storm Surge and Extreme Water Levels	Again, more space is dedicated to the aspect of results with most uncertainty. This can mislead. The use of GFDL-ESM2D needs to come with a health-warnign signpost: because a high-end analysis is valuable, and because of the addional ucertainty therein, we have to use a 5 parameter statistical model (which introduces further degrees of freedom).	Prof Kevin Horsburgh	Regarding space: I spent a long time wording this section so I'm nervous of introducing ambiguity by rewording. It is more complicated to describe the situation with the 5-parameter model and I found that took more words. Regarding the GFDL-ESM2M ananysis: the reason for using a 5-parameter model is not "additional uncertainty" but a spatially coherent signal in the 5th parameter, as stated at the end of section A1.4 (page 137, line 22 onwards in the draft as sent out for peer review).	
96	3.3 Projections of Changes in Wave Climate	You should explain why it was not possible to use GFDL-ESM2D in the wave projections. This is a consistency issue. If that model is worthy of focus in the surge section why is it not used here? The surge/wave sections should not appear disjointed. They both depend on strong winds so why treat one property differently to the other	Prof Kevin Horsburgh	The waves work was opportunistic and was completed with limited resource. In much the same way that the surge work used the Euro-Cordex data, the waves work used whatever was available from the COWCLIP project. At the time this did not include GFDL-ESM2M. In recognition of the limitations we have had on driving data, we have added a caveat in section 1 of the report. We have also added text in the Caveats / Limitations section at the start of the Marine Report to discuss the limitations of model data availability for the Storm Surge and Waves work.	
103	4.3 Potential Changes in Tide and Surge Characteristics	There are better known oceanographic tidal metrics (e.g. mean srping range, etc); probably a bit late to use any of those	Prof Kevin Horsburgh	Accepted. We chose standard deviation due to time constraints. [TH]	

105	4.3 Potential Changes in Tide and Surge Characteristics	The consistent result between this work and ours is that the variability of the weather is the crucial thing. Centennial varibaility of storminess is the dominant physical factor	Prof Kevin Horsburgh	That is not quite the point we were trying to make here. I think you found that skew surge is more-or-less independent of the high tide level (?) We found that skew surge is more-or-less unaffected by a change in the mean water level. So although these findings are not identical they do have a common factor of skew surge being unaffected by underlying water level, which is why we say they are "not inconsistent". This particular section (the case studies, 4.3.2) does not reveal anything about centennial variability of storminess. [TH]	
111	A1.3 Storm Surge Modelling	I assume the surge model (CS3) still does its normal internal temporal interpolation between the 6 hourly wind data it receives. Otherwise the wind forcing is stepwise. So the winds are interpolated in time.	Prof Kevin Horsburgh	The winds are not interpolated during the six hours between the available RCA4 fields. This is a modelling choice intended to maintain realistic maximum winds at each timestep, but we acknowledge that alternative approaches may be equally valid (or better). We note again that evaluation against observations (and our behind-the-scenes inspection of snapshots and movie loops of atmospheric forcing and sea surface elevation response) suggests that the model is doing something sensible, and we reiterate that we think it unlikely that this choice would be able to mask (or spuriously create) a significant century-scale change. We find no evidence of spurious surges associated with the six-hourly forcing (see https://drtomhoward.files.wordpress.com/2018/07/no_spurious_6_hour_surges4.pdf) [TH]	

Land report

ID	Section	Comment	Reviewer initials	MO response	Post Moderation Meeting Met Office comments
1	General	I have read and will comment on only the Sections relevant to Land Strand 2, i.e. Sections, 1, 3, 5. My first general comment refers to Section 3. I think that the high climate sensitivity of GC3.05 should be discussed up-front in an open and confident manner. Here it is not referred to until Section 3.5 and then only in a few lines on p 56 lines 17-22. In addition, the spread of the GC3.05-PPE is not large enough to encompass a suitable range of outcomes (with the majority of it coming from range of GHG scenarios). These reasons, particularly the first, are why it was decided to add as a major component in Strand 2 an ensemble of some CMIP5 models. My second general comment is that many of the Figures for Sections 3 and 5 are not framed in a way that is consistent with the storyline nature of Strand 2 and the way that it is envisaged that users should approach it. GC3.05-PPE and CMIP5-13 ensemble means, medians or lowest or highest 10 percentiles should not be considered or shown. The ensembles are not set up for this, and it is misleading. Instead either values for all the ensemble members should be shown, or maps for the member that has the lowest, median or highest value in some quantity should be shown. I assume that the same comment applies to Strand 3 results. For example, Fig. 3.8a could show the geographical distribution of bias for the GC3.05-PPE member with the lowest mean average bias, for the member with the median average bias and the member with the highest average bias. Fig. 5.2a could show the distribution for the GC3.05-PPE member with the lowest average change in surface T over England or over the UK, the member with the median average change and the member with the highest average change. It can then show the same for the CMIP5-13 ensemble and for the RCM-PPE.	Prof Sir Brian Hoskins	These comments are repeated separately below, as #49 and #50. See there for our responses.	
2	1.2 User and science drivers for UKCP18	Also mention features that are not so good if any	Prof Sir Brian Hoskins	Williams et al (2018) didn't report any specific worsening of biases in GC3.1 relative to HadGEM2-AO, but several characteristic biases do remain present in GC3.1. We added the sentence: "However, several long-standing systematic biases remain, including excessive global average precipitation,cool and warm sea surface temperature biases in the northern hemisphere and Southern Ocean respectively, and insufficient frequency of blocking anticyclones in the northern hemisphere."	
3	1.3 Three strands of UKCP18 land projections	in the first paragraph,weaknesses of the PPE approach, e.g. at best pdfs in model world, should be mentioned, different fields and locations not related. This gives a better basis for doing the Strand 2 approach.	Prof Sir Brian Hoskins	We have amended the second paragraph of section 1.3 to read: "In the light of the user drivers of section 1.2, the main limitations of Strand 1 are that the large sample of potential outcomes needed to support probabilistic projections is not suitable for derivation of a small set of storylines, that results are only provided for a limited set of climate variables, and that the projections lack the full spatial and temporal coherence of raw climate model output. The latter feature is due to the need to apply various statistical techniques within the methodology (see section 2). These issues are addressed through provision of Strands 2 and 3, which are designed to respond to user drivers 2 and 3 for a limited set of plausible realisations capable of supporting a wide range of impacts studies and development of narratives." We note that the Strand 1 projections, while not fully coherent in time and space, are far from completely unrelated in this sense (see response to #40). Also, we think it is misleading to single out Strand 1 as existing "in model world". This is because Strands 2 and 3 are also derived from climate model output, and are therefore subject to the same caveats relating to common biases, and being limited to reflecting the knowledge built into the models. See response to #95.	
5	1.3 Three strands of UKCP18 land projections	I think it would be preferabl e as far as possible to refer throughout to GC3.05-PPE, rather than shortening it to PPE	Prof Sir Brian Hoskins	Done.	
8	1.4 Emissions scenarios	replace "lacking" with "without"	Prof Sir Brian Hoskins	Done.	
11	3.2 The new climate model underpinning the Strand 2 PPE	"some" makes it sound like an interim version of the model. Hopefully we can say "most".	Prof Sir Brian Hoskins	Changed from "some improvements " to "the main improvements" (in terms of their physical impact). We have also expanded Appendix D to list the full set of differences between GC3.05 and GC3.1.	
12	3.2 The new climate model underpinning the Strand 2 PPE	mention some things that are not improved	Prof Sir Brian Hoskins	In most regions, GC3.1 does improve SST biases, relative to HadGEM2-AO. Added the following sentence: "Worldwide, SST biases are generally reduced in GC3.1 compared to HadGEM2-AO, although the warm bias in the Southern Ocean is slightly larger (Williams et al., 2018)."	

16	3.3 Design of ensemble projections	In the paragraph finishing here there are a lot of seemingly rather arbitray choices made, e.g. pre-industrial forcing, initial SST & SSS, seasonal cycle, relaxation time. More critical discussion is needed	Prof Sir Brian Hoskins	Reworded to clarify the motivations for these choices: "The calibration phase involved only the standard PPE variant (STD). Since the subsequent transient simulations required initial conditions typical of the beginning of the 20th century, we used constant pre-industrial values for greenhouse gas concentrations and aerosol emissions taken from standard CMIP5 data (https://cmip.llnl.gov/cmip5/forcing.html). As observational coverage of sub-surface temperatures and salinities was sparse around 1900, the simulation was started from an ocean state derived from temperature and salinity observations for 2004-8 (Ingleby and Huddleston, 2007).During this simulation, sea surface temperatures and salinities were continuously and linearly relaxed towards a seasonal cycle of prescribed spatial fields representing observed conditions typical of late 19th and early 20th centuries (Rayner et al., 2003; Good et al., 2013), while sub-surface ocean conditions adjusted towards the applied trace gas, aerosol and surface forcing." We also added a footnote: " A seasonal cycle of prescribed sea surface temperatures and salinities was applied, in order to facilitate the subsequent calculation of seasonally-varying flux adjustments that would limit the development of biases in these variables on a seasonal basis."	
17	3.3 Design of ensemble projections	A proper discussion is required on the reason for including flux adjustments and some critical commentary given - e.g. the hypothesis that they are OK to include unchanged in future.	Prof Sir Brian Hoskins	Added: "The main benefit of applying flux adjustments is that they reduce the potential for biases in SST to reduce the credibility of projected regional climate changes (He and Soden, 2016), with the caveat that use of flux adjustments can also affect aspects of the ocean circulation that are sensitive to variability in surface energy or buoyancy forcing, such as the AMOC (Collins et al., 2006). Also, holding the flux adjustments constant from year to year involves the key assumption that the sources of bias that they are intended to counter are invariant under climate change."	
18	3.3 Design of ensemble projections	How about changes to/drifts in aspects of the ocean circulation?	Prof Sir Brian Hoskins	We see internal variability of a few Sv in AMOC during this run, but no clear sign of a long-term drift.	
19	3.3 Design of ensemble projections	Re the footnote: any estimate of the impact of this?	Prof Sir Brian Hoskins	Added to footnote: "The effects of the change were tested. It led to changes in the detailed patterns of runoff over land, but no significant impact on the ocean surface freshwater budget at basin scales."	
20	3.3 Design of ensemble projections	Any idea of the origin of these instabilities? Did they occur at all in the other runs?	Prof Sir Brian Hoskins	In one case the crashes were related to atmosphere model instabilities associated with the simulation of snow cover on the Tibetan Plateau. In the other case, the source was also atmosphere model instabilities, but the precise cause was not identified. Other members also crashed occasionally, but with nowhere near the same frequency.	
22	3.4 Evaluation of combined Strand 2 ensemble	This is a major point - for the CMIP5-13 models the drift was in general somewhat lower than in the GC3.05-PPE and flux adjustment was not deemed necessary	Prof Sir Brian Hoskins	Here, we are simply pointing out that the reduced SST biases in the PPE runs is an expected consequence of the application of flux adjustments, and does not necessarily indicate a fundamental improvement in the simulation of ocean transports or surface fluxes. Note also that we don't say (or know) whether climate drift in the PPE runs is typically smaller or larger than in the CMIP5 runs. The fact that flux adjustments were not used in CMIP5 runs does not mean that using them in the PPE to restrict regional SST biases might not be beneficial - in particular, by reducing the effects of SST-related biases on projected changes in terrestrial climate (e.g. He and Soden (2016), cited in the report). We have revised the relevant sentence to make our point more explicit: "This is likely to be due, at least in part, to the use in the latter of flux adjustments, and should not be taken as an indicator that GC3.05-PPE necessarily simulates better than CMIP5-13 the processes giving rise to regional values of SST."	
23	3.4 Evaluation of combined Strand 2 ensemble	The NAO results discussed in this paragraph are quite concerning given its importance for UK weather/climate. Can any more be said about it?	Prof Sir Brian Hoskins	The bias in *interannual* variability in NAO is present in both the CMIP5-13 and GC3.05-PPE ensembles, and varies with ensemble member. We think the text captures that point OK. Interestingly, the *intraseasonal* variability (as measured by attribution of days to weather types) is captured quite well. We have amended the text to cross-reference these contrasting results on different time scales. We've also revised some ambiguous wording from the original draft: The uncertainty ranges in Fig 3.9 show variations through the 20th century in the strength of *interannual* variability, but not variability in the phase of the NAO on multidecadal time scales (as might be revealed by low-pass filtering the data prior to calculation of variance statistics). We've added a footnote to this effect. Inspection of the NAO time series suggests that observations show more variation on 20-30 year time scales than the models do, though it is of course challenging to disentangle genuine internal variability from potential influences of extenal forcing. In the models, we think that the (forthcoming) availability of multidecadal control simulations will help with this, and that analysis of low frequency variability should be deferred to that point. In terms of potential effects on stakeholder analysis of UK impacts, we have added this comment: "Therefore,the effects in impacts assessments of biases in simulated characteristics ofthe NAO are likely to vary according to the time scales of interest in different applications: Biases can be quite different on intraseasonal,interannual or interdecadal time scales, and will need to be evaluated on a case-by-case basis. "	
24	3.4 Evaluation of combined Strand 2 ensemble	However it should be recognised that the large aerosol cooling could mask an overactive response to GHGs in the 20th century, and that this masking will occur to a much smaller extent in the projections	Prof Sir Brian Hoskins	Added: "For example, the strong aerosol cooling may have masked the response to greenhouse gas forcing during the 20th century, but such masking would reduce substantially during the 21st century, as aerosol emissions are assumed to reduce in the RCP8.5 scenario. "	
25	3.4 Evaluation of combined Strand 2 ensemble	The analysis in this section is thorough, but the results for Europe are disturbingly bad. Also, please note my general comment that individual ensemble members must be shown in all figures and discussed in the text, no top 10 percentiles etc. This will mean a significant change in much of this section.	Prof Sir Brian Hoskins	We assume this comment relates specifically to the PPE. The cold bias in GC3.05-PPE is a significant European bias, and winter preciptiation is too high (noting the caveat around observational undercatch - see #67), as is the average NAO amplitude. However, in other aspects the GC3.05-PPE performs competively with CMIP5-13 , so we think it's a mixed picture. Regarding the comment about presentation, see #50.	
27	3.5 Projected global changes in combined Strand 2 ensemble	This means that the PPE approach is not working as well as hoped. Indeed, most of spread is from GHG input.	Prof Sir Brian Hoskins	See #31. We agree. There is (relative to GMST) typically more spread in GC3.05-PPE at regional scales, but we were hoping for more. That said, it remains the case that combining the two enembles does lead to a broader set of UK realisations than considering CMIP5-13 in isolation (Fig. 5.3). Nevertheless, understanding the relative lack of spread in GC3.05-PPE will be a future priority, to inform how the next generation of UK projections should be constituted.	
28	3.5 Projected global changes in combined Strand 2 ensemble	ECS is used here but I cannot see that it has been introduced/explained. The high ECS of GC3.05 is a crucial point and is just slipped in here. Please see my general comment.	Prof Sir Brian Hoskins	ECS is defined earlier, in section 2. However, we have now redefined in section 3.5, for convenience. For the broader issue regarding the GC3.05-PPE warming/ECS, see #49.	
29	3.5 Projected global changes in combined Strand 2 ensemble	Is the difference in the value of CRF in GC3.1 quantitatively consistent with this being the/a major factor in the difference in its ECS?	Prof Sir Brian Hoskins	Yes.	

30	3.5 Projected global changes in combined Strand 2 ensemble	A very weak statement for all the work involved!	Prof Sir Brian Hoskins	We think it's an accurate summary of the figure, but yes, in terms of spread we were hoping for more examples where combining the two ensembles would lead to broader sampling of patterns of change. When full (non-normalised) transient changes are considered (as in Fig. 3.22), we do see more examples of a broader combined spread, due to the influence of the high global warming in GC3.05-PPE.	
31	3.5 Projected global changes in combined Strand 2 ensemble	Again the PPE method gives disappointing results in this case, cf Murphy et al	Prof Sir Brian Hoskins	We agree. The degree of spread does vary, but overall the new PPE is less effective than HadCM3-based PPEs were, in generating diversity. It will be very important for future work to understand what controls, or inhibits, the spread we hoped to see.	
33	3.5 Projected global changes in combined Strand 2 ensemble	This is a very complicated example of changes with time of normalised warming pattern	Prof Sir Brian Hoskins	We assume this comment applies to Fig. 3.22, which is an example of full transient changes, not just the normalised pattern contribution. Reworded to clarify.	
36	5.1 Comparison of projections from Strands 1-3	This section needs considerable change as per my geberal comment about how to summarise and compare the Strand2 results. I cannot list all the places this is relevant as it pervades the section.	Prof Sir Brian Hoskins	See response to #50. Although we do not think it would be wise to *replace* the 10-50-90 percentile panels in Fig. 5.2 with maps showing individual low, medium and high members, *adding* an example of the latter is an option we'd feel more comfortable with. We have produced examples, that could be discussed during the September 2018 moderation meeting. As also noted in #50, we are also not in favour of adding a large compendium of figures showing each individual member in the report (though placing these on the UKCP18 website is an option). We are concerned about making the report too long, and overwhelming readers with information that cannot (on the tight timescale of UKCP18 launch) be supported by anything more than superficial discussion.	Maps showing individual examples of members with low, central and high UK average responses now added.
37	5.1 Comparison of projections from Strands 1-3	This gets tricky as all the LS2 storylines are all seen as possible. It is only unlikely viewed with the LS1 methodology. Probability should not be attached to LS2 results.	Prof Sir Brian Hoskins	We have changed the sentence to read: "This indicates that the PPE responses in Strand 2 should form a useful source of information for studies of heat-related impacts in summer climate that are plausible, but lie near the upper end of the uncertainty ranges provided by Strand 1." However, this is worth discussing. "Unlikely but plausible" is used as a way of describing H++ scenarios, for example, and there is not necessarily a contradiction between the two terms. Ways also need to be found to communicate the point that saying where a Strand 2 simulation lies within the Strand 1 range does not mean we are attaching low credibility to the particular realisation in question. Rather, we are just pointing out that its projected outcome lies towards the upper end of outcomes implied by a broader collection of evidence.	
38	5.1 Comparison of projections from Strands 1-3	Fig. 5.7 on storm-track changes is a more difficult one to see how to change. It could be done by showing changes in the numbers of storms 40-60N and 60-80N for each member.	Prof Sir Brian Hoskins	See response #50. We have revised this figure to improve the presentation (see #91), but we actually see this as a good example of the benefits of showing ensemble-averages. It is interesting that GC3.05-PPE and CMIP5-13 show similar patterns of change, and the fact that these similarities emerge from two ensembles of totally different construction is a simple message that we can illustrate immediately, with a straightforward plot. This is not, of course, to say, that the spread of changes isn't interesting too, and indeed these could be placed on the project website. However, we do not see it as practical (given the project timescale and the (assumed) need to keep the report to a reasonable length), to make the Science Report a place where we discuss the new realisations at an individual level. Rather, we are seeking to provide some focused summary information that provides a baseline for research and applications scientists to build on in the future. Investigating and understanding the storm track realisations individually would be a good example of work that we hope will follow.	
39	5 Projections of future variability and change for the UK	This section appears to be written from an LS1 perspective	Prof Sir Brian Hoskins	Section 5.1 does begin by focusing on uncertainty ranges (in Figs. 5.1 and 5.2), because we think it's important to frame the new realisations in the context of the risk information provided by Strand 1. But once the discussion moves to time series of annual anomalies (Fig. 5.3) and joint changes relevant to impacts (Fig. 5.5), we place more emphasis on how the realisations can be used to look at changing risks of extremes, examples of future drought, detailed local events driven by specific type of synoptic event in case studies, etc. So, we believe that the storyline and impacts aspects, where Strands 2 and 3 come into their own, are also well covered. That is certainly the intention, at any rate !	
40	5.2 Interpretation and use of the projections	Need to give caveats e.g. about model world & lack of spatially coherent data	Prof Sir Brian Hoskins	We have added the following sentence: "However, results are only available for a limited set of key climate variables (Fung et al., 2018), and the realisations do not possess the full spatial and temporal coherence available directly from climate model output, due to the extensive statistical processing required to create the results." For the record, we note that whilst the Strand 1 results lack the *full* spatial coherence of raw climate model output, they still possess a significant degree of coherence - the are far from being spatially incoherent ! This is partly because they are derived from climate model output, and also because we use statistical emulation methods designed to build in as much coherence as we can. We will discuss this more in the specialist journal paper that will follow during 2019. We have also revised the description of Strand 1 to avoid the "robust as possible" phrase, and express its status as summarising known uncertainties incorporated in models. See response to #99.	
41	5.2 Interpretation and use of the projections	different rather than simpler - examples of physically plausible climate outcomes	Prof Sir Brian Hoskins	Changed to: "Strands 2 and 3 were developed primarily to support user requirements for more flexible datasets that provide specific examples of physically plausible climate outcomes..."	
42	5.2 Interpretation and use of the projections	Not percentiles for Strand 2 (and 3)	Prof Sir Brian Hoskins	We have revised this piece of text to emphasise the distinction between the Strand 1 probabilities and Strand 2-3 percentile ranges (see response to #50). The paragraph below does go on to discuss usage of Strand 2/3 results in terms of using specific realisations to support the uncertainty analysis of Strand 1, consistent with the storyline/non-probabilistic nature of Strands 2/3.	
43	5.2 Interpretation and use of the projections	"unlikely" is an LS1 perspective. Surely here we should be emphasising that they are projections that are physically plausible and come from a later model with new science in it.	Prof Sir Brian Hoskins	Unlikely is indeed a Strand 1 perspective, but providing estimates of uncertainty and relative likelihood is the purpose of Strand 1, so we think this perspective is important. We agree that we should (and hopefully do) emphasise the plausibility of gthe new simulations, but as you and other reviewers point our elsewhere, those runs have their own biases and limitations, so we don't think it would be wise to appear to be selling them at the expense of Strand 1, simply because they are new.	
44	5.2 Interpretation and use of the projections	Replace "current" by "recent models and the physical knowledege incorporated in them"	Prof Sir Brian Hoskins	Done.	
45	5.2 Interpretation and use of the projections	Somewhere around here we should mention data set size increases with resolution.	Prof Sir Brian Hoskins	Noted in text.	
47	5.2 Interpretation and use of the projections	The changed figures and text I am suggesting would link directly with this important usage.	Prof Sir Brian Hoskins	See #50	

48	1 Introduction	This is a general comment on the Land Report. I have read and will comment on only the Sections relevant to Land Strand 2, i.e. Sections, 1, 3, 5.	Prof Sir Brian Hoskins	Noted.	
49	1 Introduction	This is a general comment on the Land Report. My first general comment refers to Section 3. I think that the high climate sensitivity of GC3.05 should be discussed up-front in an open and confident manner. Here it is not referred to until Section 3.5 and then only in a few lines on p 56 lines 17-22. In addition, the spread of the GC3.05-PPE is not large enough to encompass a suitable range of outcomes (with the majority of it coming from range of GHG scenarios). These reasons, particularly the first, are why it was decided to add as a major component in Strand 2 an ensemble of some CMIP5 models.	Prof Sir Brian Hoskins	Our suggestion to add CMIP5 models to Strand 2 was motivated by a general expectation that including structural model diversity would broaden the range of realisations. This motivation related to regional *patterns* of change (for example, without CMIP5 models, Strand 2 would not include any realisations in which UK summer rainfall does not dry in the future), as well as global temperature outcomes. At the time of decision, we did have information from idealised experiments suggesting the possibility of some high-end global warming outcomes, but we didn't yet have the transient GC3.05-PPE simulations. Therefore, we don't think it would be appropriate to discuss the high warming earlier, in the *experimental design* part of section 3, as if it was the sole (or dominant) reason for adding the CMIP5-13 models. This is why we introduced the discussion of high warming in section 3.5, rather than earlier in section 3. We do, however, agree that the high warming result should be mentioned earlier in the report. We will address this by adding an Executive Summary at the front of the document, as suggested in comment #230. In this report, we are happy to include open and confident discussion that we can support with peer-reviewed evidence. However, we are constrained by the newness of the GC3.05 simulations and the project delivery timescale. We think it would be unwise to include "work in progress" information, due to the risk that some of this might later have to be retracted. In particular, we do not agree that the UKCP18 land science report is the right document in which to provide an extensive discussion of ECS. This is motivated by three reasons and follows discussion with the Met Office Chief Scientist on the issue of presenting ECS: • Work on understanding the ECS has not used GC3.05. Rather, it has used GC3.1 and UKESM1. • Aside from the standard model variant, GC3.05-PPE consists of members that each have different ECS values, but these values cannot be estimated from the scenario experiments available. • Work on understanding ECS in GC3.1 is not yet peer reviewed or published and is still evolving. Presenting a non-peer reviewed version risks being out of date and incomplete. The work will be published in due course and is relevant to multiple Met Office projects, including our CMIP6 contribution. However, we do agree with the need to address the high warming issue in the UKCP report and show how the PPE compares to CMIP5 models. We propose to do this by showing a comparison of the RCP8.5 warming response in the GC3.05 PPE with the IPCC AR5 simulations. This is a more appropriate comparison because it uses the scenarios actually applied in UKCP18. This marks a change from the 'soft caps' in which the ECS estimate of 6.2K (for GC3.1) was included. We agreed	We added the following text, to give a sense of where ECS values for GC3.03-PPE may lie with respect to the current IPCC assessment: "In AR5, IPCC assessed ECS to have a likely range of 1.5-4.5ºC (Collins et al., 2013), and also judged thatthere is a small probability (of up to 10%) that ECS exceeds 6ºC. The levels of 21st century warming simulated for the RCP8.5 scenario suggest that most of the GC3.05-PPE members are likely to possess ECS values above 4.5ºC. ". At the Moderation Meeting, the PRP advised us not to add an Executive Summary, so we have not done this.
50	1 Introduction	This is a general comment on the Land Report. My second general comment is that many of the Figures for Sections 3 and 5 are not framed in a way that is consistent with the storyline nature of Strand 2 and the way that it is envisaged that users should approach it. GC3.05-PPE and CMIP5-13 ensemble means, medians or lowest or highest 10 percentiles should not be considered or shown. The ensembles are not set up for this, and it is misleading. Instead either values for all the ensemble members should be shown, or maps for the member that has the lowest, median or highest value in some quantity should be shown. I assume that the same comment applies to Strand 3 results. For example, Fig. 3.8a could show the geographical distribution of bias for the GC3.05-PPE member with the lowest mean average bias, for the member with the median average bias and the member with the highest average bias. Fig. 5.2a could show the distribution for the GC3.05-PPE member with the lowest average change in surface T over England or over the UK, the member with the median average change and the member with the highest average change. It can then show the same for the CMIP5-13 ensemble and for the RCM-PPE.	Prof Sir Brian Hoskins	We understand these concerns over potential misinterpretation. We agree that since Strands 2 and 3 are designed for storylines (and flexible assessments of impacts) rather than quantification of uncertainties, their results should not be interpreted probabilistically. However, we see the purpose of this report as providing some basic context to underpin *subsequent* use of Strands 2 and 3 in this way, rather than jumping straight to a storyline-based style of presentation. This context, in our view, should include illustrations of systematic model biases in the evaluation sections, and also some basic information of how the envelopes of future change given by Strands 2 and 3 compare with the broader representation of uncertainties in Strand 1. Given this purpose, we think that ensemble-means or medians are appropriate to show, as they give (by construction) a more reliable estimate of the spatial distribution of the systematic component of model error, or of the mid-point of some range of future changes, than would an individual realisation that lay "in the middle" in some spatial average sense. Similarly, maps of the 10th and 90th percentiles give estimates of the span of changes that are locally more reliable than a plot showing single realisations. We think, therefore, that if we were to replace the original figures with figures based on single realisations, this could create more risks of misinterpretation than it resolves. We also note that ensemble means and ranges are often plotted to characterise non-probabilistic results from multimodel ensembles or PPEs in journal papers, as well as by IPCC. Therefore, we don't agree that the limited sizes of the relevant ensembles, or their design criteria, necessarily disqualify such metrics from use. However, we certainly agree that provision of appropriate caveats is important. Some text of this nature was provided in the original draft. We have revised this (particularly in section 5.1, when introducing Figs. 5.2a-d) to spell out more strongly the difference between the results of Strands 2 and 3 cf Strand 1. The Strand 1 results are now called "probability levels", and the Strand 2 and 3 percentiles are discussed in terms of frequency within the relevant ensemble. The purpose of the comparison (basic framing of ranges of response) is also stated more explicitly, and a warning about not over-interpreting the significance of the Strand 2/3 median is included. The figures have also been revised, with wording changes to emphasise the difference between the probabilistic projections and the realisations, supported by addition of white space. Maps showing responses of each Strand 2 member will be available from the user interface. We have included a link to this, as part of the revised discussion outlined above. We would prefer not to get into a specific discussion of the individual responses in the Science report itself, as this would inevitably be of a "show and tell" nature (due to the newness of the Strand 2 and 3 products), which would increase the (already high) ratio of information to understanding in this long report. The philosophy behind Strands 2 and 3 is that the UKCP18 report should provide a general evaluation of credibility using standard model assessment metrics of broad interest and relevance, and that more detailed understanding of application-specific capabilities and limitations will then emerge in slower time, as the research and user communities analyse the results and develop deeper insights.	Following the Moderation Meeting, it was agreed that the plots comparing local ranges would be retained, but using results from specific members close to the relevant percentiles of the underlying frequency distribution of outcomes, rather than presenting limits labelled as specific percentiles. We also added plots showing examples of individual members with low, central and high UK average responses, to emphasise the storyline nature of Strands 2 and 3. this has been done, with text revised accordingly.
51	1 Introduction	General comment: The design of the UKCP18 methodology is an impressive and outstanding piece of work. I think it will be remembered for a long time as the most comprehensive approach to construct probabilistic climate projections. I have also been impressed by the write-up, which obviously had to happen under considerable time pressure and is neverthelss of overall very high quality. On the other hand, I think that the model versions used are not fully convincing. The GC3.05 simulations have a limited correspondance to observations in the last 100 years. In retrospect I wonder whether making more thorough use of the CMIP5 projections would not lead to a more convincing uncertainty assessment.	Prof Dr Christoph Schär	We appreciate the recognition of our efforts, and of the time pressure that we have been under. In fact, it is Strand 1 that is the primary source of information on uncertainties, as explained in sections 5.1 and 5.2. These are constructed from HadCM3-based PPEs that do provide uncertainty ranges competitive with those of CMIP5, and the main strength of Strand 1 as an uncertainty product stems from the combination of their results. Strand 2 is designed to provide a limited set of realisations. Diversity is important in these too, and it is true that the ranges provided by GC3.05-PPE are narrower than those of CMIP5, notably for GMST, though the contrast in spread is less marked for the UK. Nevertheless,an important result is that adding the GC3.05-PPE runs to CMIP5 results gives a broader set of realisations than would be obtained from the screened set of CMIP5 models in isolation. Also, we assess the *overall* skill of the PPE runs as competitive with CMIP5, while acknowledging that GC3.05 does contain some notable biases. A balancing consideration is that other aspects of the GC3.05-PPE runs (including the North Atlantic storm track) are comparatively well simulated. for these reasons, we do not think that reducing the cohort of GC3.05 runs to accommodate more CMIP5 models would be an overall benefit, since any gains in diversity would be offset by reductions in plausibility.	

52	1.2 User and science drivers for UKCP18	Please discuss in more detail how GC3.05 differs from GC3.1	Prof Dr Christoph Schär	A discussion of the full set of differences has been added to Appendix D.	
53	1.4 Emissions scenarios	... run with presribed GHG AND AEROSOL concentrations?	Prof Dr Christoph Schär	It's more complicated than that, so not possible to describe in a simple phrase. The Strand 2 PPE simulations used prescribed aerosol precursor emissions with an interactive scheme. Different CMIP5 runs used different aerosol schemes, with differing levels of complexity. The Strand 3 runs used prescribed changes in aerosol optical and cloud properties, derived from Strand 2 PPE output. This is described in sections 3.3e and 4.2. Here, we are making a specific top-level point about how CO2 is treated, so we think it better to stick with this, and leave the detail on treatment of other species to the detailed methodology sections. However, we do refer briefly to aerosol forcing at the bottom of this paragraph relevant para, so their role in emissions scenarios and related uncertainties does get mentioned in this part of the text.	
54	2.1 Overview	To what extent is the use of an earlier HadCM version consistent with the use of GC3.05? This should be discussed in light of significant differences in physics (e.g. pertaining to the aerosol moduless used in HadCM3 and GC3.05).	Prof Dr Christoph Schär	This is discussed in section 5.1: "The HadCM3- and GC3.05-based PPEs were both developed from UK climate models. Their land surface components are structurally similar, however the two models contain many structural differences, including their representations of atmospheric dynamics, large-scale cloud and boundary layer processes (Walters et al., 2017). The overall ranges of response seen in Strands 1 and 2 for GMST (Fig.3.19), and patterns of response (Fig. 5.1), are therefore a consequence of combining three quasi-independent lines of modelling evidence, namely: HadCM3-based PPEs (Strand 1), GC3.05-PPE (Strand 2) and CMIP5 multi-model ensembles (Strands 1 and 2, via emissions- and concentration-driven experiments respectively)."	
56	2.2 Technical summary of methodology	Major comment: How do the estimates of ECS compare against the international literature? In particular, how does it compare against IPCC AR5 (Chapter 12)? In the AR5, a 66% probability has been stated of the ECS to be between 1.5 and 4.5K, a <1/10 chance to be below 1 K, and a <1/20 probability that the ECS is above 6K. The current estimate appears to be much narrower than that of the AR5. Can this be justified?	Prof Dr Christoph Schär	This comment refers to discussion of Figure 2.3. Since the pdf of ECS in this figure is a prior distribution (see response to #220), it would not be appropriate to compare it to assessments in the international literature that include consideration of observational constraints. It is included in Fig. 2.3 to illustrate methodology - ECS is a control variable for the simple climate model that we use to sample transient responses from time-invariant metrics of climate change. Doing such a comparison for the *posterior* distribution of ECS would be very interesting, and we agree that it should be done. However, this will require significant supporting analysis and discussion to allow readers to understand the causes and implications of any differences cf, say, the AR5 assessment. It is not straightforward, for example, to pick apart whether a narrower distribution signifies a better use of constraints, or a less complete consideration of evidence. We think, therefore, that this comparison would be better placed in the specialist journal paper(s) on Strand 1 that will follow launch of UKCP18. Instead, we have provided a comparison against AR5 of projected transient global temperature outcomes from Strand 1, for RCP scenarios (Table 2.2 in revised version - see comment #233). We think this addresses the spirit of this comment, and is of more direct relevance to UKCP18, since it is the transient global changes that drive changes in UK variables.	
58	2.2 Technical summary of methodology	It appears that a full comparison against CMIP5 is missing (both for global and regional transient projections). Such a comparison could be shown along the bottom panels of Fig.2.4a and Fig.2.4b. It appears that the RCP8.5 projections of the current report are significantly warmer than those of CMIP5. This difference should be discussed.	Prof Dr Christoph Schär	Global and regional comparisons against our CMIP5-13 set of simulations are presented in Figs. 3.19, 5.3 and 5.4. We don't think such a comparison should be added to Figs. 2.4a,b, as the purpose of those figures is to illustrate methodology, rather than develop take-home messages about the projected ranges. We have, however, added a presentation of ranges of GMST response to section 2.3, compared against the AR5 ranges derived from CMIP5 (#233). The upper ends are a bit higher in UKCP18 Strand 1(partly due to sampling carbon cycle feedback uncertainties), but overall the Strand 1 vs AR5 ranges are quite similar.	
59	2.3 Examples of the probabilistic projections	Overall the growth in uncertainty is significant. For SE England, there is a 25% chance that there is no reduction in JJA precipitation by 2085 (and a 10% chance that JJA precip is increasing).	Prof Dr Christoph Schär	See #231. Figure 2.6 will be replaced by a new figure showing an improved precipitation method, which reduces the wet tail seen in the original version.	New Figure included. The improved calculation shows a smaller chance of increasing precip in SE England in the late 21st century, and an uncertainty range more similar to that of UKCP09.
60	2.3 Examples of the probabilistic projections	The placeholder text is important. (1) Is it possible that the uncertainty in UKCP18 grows faster, as overall the overlap between the transient GMST projections and past observations is smaller? (2) What does this imply for other national scenarios in Europe (which provide smaller uncertainty ranges than the current UKCP18).	Prof Dr Christoph Schär	See #231. Figure 2.6 will be replaced by a new figure showing the improved precipitation method, in the final version.	See #59 above.
61	2.3 Examples of the probabilistic projections	Based on the material presented, I disagree with this interpretation, at least for the JJA TAS signal. You should also explore another hypothesis, related to the fact that GC3.05 does not capture the past warming. According to Fig.3.11, the GC3.05 PPE mean GMST deviates strongly from HadCRUT4 in terms of GMST in the period 1950-2010. Likely this also affects the changes in the two UK variables shown in Fig.2.9, i.e. there is an underestimation of the observed changes in GC3.05 PPE mean.	Prof Dr Christoph Schär	Figure 2.9 is a Strand 1 result, derived from a combination of HadCM3-based PPE results and CMIP5 ESMs. The GC3.05 runs are not included in Strand 1, so the results of Fig. 3.11 are not relevant to Fig. 2.9.	
63	3.2 The new climate model underpinning the Strand 2 PPE	Major comment: Here a more detailed discussion is needed. How does the current HadGEM3 model compare in its aerosol treatment against the version studied by Malavelle et al (2017). In this study a comparison was presented between observations and simulations in the event of an Icelandic volcanic eruption. It was found that aerosol effects were strongly overestimated by some HadGEM3 version. Ref: Malavelle, F. F. and et al, 2017: Strong constraints on aerosol–cloud interactions from volcanic eruptions. Nature, doi: 10.1038/nature22974	Prof Dr Christoph Schär	We included a paragraph on the Malavelle et al paper in section 3.4, following the discussion of the historical temperature record (Fig. 3.11). In fact, they used a vesion of the HadGEM3 atmosphere model close to that of GC3.05. This was coupled to two aerosol schemes, one being the GLOMAP-mode module used in GC3.05, the other the previous CLASSIC scheme used in HadGEM2. It was the CLASSIC experiment that overestimated the indirect effects, whereas the simulation relevant to Strand 2 performed well. This doesn't mean that the total aerosol forcing in GC3.05 isn't too strong, but it does illustrate the risks of over-interpreting what simple global-average metrics do or don't tell us about the credibility of aerosol processes in models. We've added a paragraph expressing these points.	

64	3.3 Design of ensemble projections	Major comments: The design of the GC3.05-PPE is a very impressive and outstanding piece of work. It is evident that objecive criteria were used as far as possible. I have nevertheless two concerns: (1) while 47 parameters were systematically varied, it appears that model switches were left untouched, i.e. effectively using one single set of parameterizations. It is a general experience in climate modeling that switching parameterizations often enables model changes that exceed those amenable with parameter changes. In particular, one may suspect that all parameter choice in essence include an overestimation of indirect aerosol effects. (2) The use of comparatively short simulations was dictated by available computing resources. In comparison to other modeling centers, probably less emphasis was put on the simulation of GMST in the historical period.	Prof Dr Christoph Schär	Thank you for for the compliment. You raise two good points. 1) When we perturbed HadCM3, there were several continuous parameters that had large effects, notably entrainment rate and ice fall speed. We also perturbed switches as well as parameters and found there was one switch to do with cloud overlap that did have a major effect. Root depth was handled as a discrete variable and that played a role, and introduction of a CAPE timescale was important for rainfall. However, other switches were less important. Entrainment rate, effect of subgrid scale orogrpahy, cloud ice and radiation parameters play a large role in the new ensemble. So it depends on the specific switches and parameters. Switches are hard to deal with because they make the sampling less efficient and make emulation harder - it takes 2^N simulations to sample N switches, maybe 6-10 times N to emulate N continuous parameters. So it was a pragmatic choice, otherwise we would have had to sample fewer parameters overall and the emulation would have become less reliable. Ideally, to better sample structural and parametric uncertainty, it would be good if perturbed ensembles were made around several CMIP5 members - e.g. Carslaw, K. S., L. A. Lee, L. A. Regayre, and J. S. Johnson (2018), Climate models are uncertain, but we can do something about it, Eos, 99, https://doi.org/10.1029/2018EO09375 2). There were three stages of filtering: 5-day, 5-year and historical performance from coupled runs. The first two stages were done quantitatively and qualitatively, and the large ensemble sizes at these stages allowed us to do the quantitative assessment. The small size of the coupled ensemble meant that the assessment becomes more qualitative, similar to our assessment of CMIP5. Qualitative assessment of coupled model output is more labour intensive, so we prioritised variables that were relevant to UK climate, or variables that required coupled model output for meaningful assessment - sea-ice, AMOC and variability of large scale indices like ENSO and NAO, for instance.	
65	3.4 Evaluation of combined Strand 2 ensemble	Fig.3.11: As this figure is using masked temperature series, there is considerable room for misinterpreation. I recommend that the same figure is shown without masking, together with observational GMST estimates.	Prof Dr Christoph Schär	This is done in Fig. 3.19, so comparisons can be made, and the differences are briefly discussed. Note, though, that creation of comparable full-field observations requires in-filling, which is an approximation. We believe that creating masked versions of the model output to match the observational coverage reduces rather than increases risks of misinterpretation.	
66	3.4 Evaluation of combined Strand 2 ensemble	Major comment: Fig. 3.11 ands 3.19: Overall GC3.05 underestimates the observed warming in the period 1900-2000 by about 0.2-0.3K, while the warming trend from 2000-2017 is overestimated by about a factor 2. The corresponding biases with CMIP5-13 are smaller. This aspect should be discussed in detail, as it could potentially become a topic of public debate. Overall it appears to me that the an overestimation of the (indirect) aerosol effect could likely play an important role (leading to an overestimation of the warming in the period prior to 2000, and an overestimation thereafter).	Prof Dr Christoph Schär	In section 3.5 we already explain the cooling and link it to the strong aerosol forcing. However, apart from the coolest member already flagged in the text, the other members lie mainly within the CMIP5-13 range, albeit mostly in the lower half. Relevant sentence changed to read : "However, during the period 1960-2000 most GC3.05-PPE members simulate long-term average values (not shown) that are cooler than the baseline period (1901-1950 in Fig. 3.11) and in contrast to observations, although they lie mostly within the CMIP5-13 range, albeit mainly at the lower end. The simulated warming rates during 2000-17 are indeed greater than observed. Some of this could certainly be due to removal of aerosols (we make this point n discussion of Fig. 3.19). Another factor could be the warming hiatus, that we also mention. We have amended the final bullet in our list of complicating factors in assesing the relationship between past and future warming. It now reads: "•The role of internal variability in the observed record (Olson et al., 2013), allied to challenges in disentangling natural and anthropogenically-forced contributions to variability on multidecadal time scales (e.g. Booth et al., 2012). For example, if the warming hiatus was mainly due to internal variability, then we would expect the observations to show less warming than most of the simulations during the relevant period, and indeed the Strand 2 runs warm faster during 1998-2014." In the longer term, aerosol-only simulations could help, though there is also interest in other follow-up experiments. See #83.	
67	3.4 Evaluation of combined Strand 2 ensemble	Observations of winter precipitation are known to systematically underestimate precipitation due to wind-driven undercatch of standard rain gauges. The error in some areas (in particular with snow fall) may well reach 30-40%.	Prof Dr Christoph Schär	Yes, we think this is an important point. We have added the following text in section 3.4b, and also refer back to it when discussing the Strand 3 results in section 4.4. "Note, however, that rain gauge observations underestimate the true level of observed precipitation due to undercatch associated with multiple factors, including the fraction of precipitation falling as snow, wind speeds and the exposure of the gauge. Biases vary with season and location, and are largest in winter, and at high elevations (e.g. Frei et al., 2003). Some authors (Kotlarski et al., 2014;Rajczak and Schär,2017) increase observed values by 20% to estimate the effects of undercatch. We do not apply such corrections in this report. However, we note that measurement biases provide an important caveat to our estimates of systematic biases in simulated precipitation (Figs. 3.12b, 3.13b, 3.14, 4.1c, 4.1d, 4.2, 4.3b and 4.4), and an important source of uncertainty in the context of bias correction methods."	
68	3.4 Evaluation of combined Strand 2 ensemble	I am surprised to see the larg positive JJA precipitation biases in the Mediterranean (amounting to 100% in some areas). It is known that many GCMs underestimate JJA precipitation together with an overestimation of temperature. Please check reasons for results. Is this due to a subset of CMIP5 models?	Prof Dr Christoph Schär	We will check, and amend the final version if/as needed.	We checked, and the results are correct, as measured relative to the E-OBS dataset. Several CMIP5 models shows these large percentage biases, though they are not so large in absolute units (mm/day), due to the small baseline valies in these regions. Also, there is significant dataset uncertainty: GPCP (used in Fig. 3.8) shows greater values than E-OBS. These points have been made in the revised text.
70	3.5 Projected global changes in combined Strand 2 ensemble	The significant differences between Strand 1 and CMIP5 on th one hand, and Strand 2 on the other hand, are a source of concern. One wonders whether assessing the uncertainty primarily bssed on Strand 2 and the associated RCM simulation is really the right decision in this situation. The systematic differences are even larger when looked at in other metrics (Fig.3.22)	Prof Dr Christoph Schär	See #238. We are not suggesting that users assess uncertainty with Strand 2 simulations, as this is the purpose of the Strand 1 probabilities. Instead Strands 2 and 3 should be seen as a source of individual realisations to provide examples of what a future UK climate might look like, or (in the case of Strand 2) to better understand the drivers of the local climate change. This needs to be reflected in the user guidance as it is in the Science Report. The advice is applicable to output from impacts models driven by Strand 2 or 3 output as well as to the model output data itself.	
71	4.1 Overview	A recent study indicates that the improvement in model climate when downscaling GCMs with RCMs is systematic and not only due to resolution. See Sørland, S.L., C. Schär, D. Lüthi and E. Kjellström, 2018: Regional climate models reduce biases of global models and project smaller European summer warming. Env. Res. Letters, 13, published online, https://doi.org/10.1088/1748-9326/aacc77	Prof Dr Christoph Schär	Added the sentence: " Regional models can also provide improvements at larger scales that might be considered well resolved by the driving GCM, especially in regions and seasons where the influence of the driving large-scale circulation is relatively weak (Sørland et al., 2018)." We would note, however, that the cited paper is a limited study involving only two RCMs, considering only the surface temperture field, and lacking a process-based explanation of the sources of the improvement found. There is also a "double-edged sword" question, which this paper ignores, of whether the existence of large-scale differences between the driving and regional models creates concerns relating to the physical consistency of regionalised scenarios. Such concerns would be particularly relevant in experiments such as EuroCordex, in cases where the global and regional model parameterisations are structurally different, and therefore give very different regional outcomes.	

72	4.1 Overview	The use of this RCM ensemble is limited as it is biased against GC3.05. In retrospect it would have been good to downscale a few of the CMIP5 members by the same RCM, to provide a broader assessment of the uncertainties using one particular RCM.	Prof Dr Christoph Schär	The decision to add a 12km ensemble of RCM simulations, driven by the Strand 2 global PPE variants, was taken relatively late in the UKCP18 project, and before the results of the global simulations were known. Given the time scale for launch of the scenarios, it was not feasible to consider adding further simulations sampling structural modelling uncertainties in Strand 3, either by driving the Met Office RCM with CMIP5 models, or by adding a screened subset of EuroCordex simulations derived from structurally-different GCM-RCM pairs. These are interesting ideas to consider as potential future additions to UKCP18. However, they will need to be tensioned against other priorities for subsequent add-ons to the package of information, in the light of available people and HPC resources. Alternative priorities would include, for example, additional global PPE simulations to understand historical biases related to aerosols, future climate feedbacks or model drift. We do include text in section 4.1 that points out EuroCordex as a potential additional resource for users wishing to consider a broader range of downscaled scenarios, so this (unavoidable) limitation in the Strand 3 design is raised in the report.	
74	4.2 The 12km regional climate model	Does this follow the EuroCORDEX protocoll and cover the respective domain?	Prof Dr Christoph Schär	The RCM-PPE runs do use the EuroCordex domain. This is pointed out in section 4.1, and we have added a reference to EuroCordex to the relevant sentence in section 4.2. The simulation periods (1981-2080 for RCM-PPE, and 1951-2100 for EuroCordex), are stated in section 4.1. Frustratingly, the forcing strategy in Eurocordex runs does not seem to be stated explicitly on the website or in papers, so we cannot comment on this. It is not clear whether consistent time-dependent changes in greenhouse gases are prescribed in the Eurocordex RCM runs (as we do in the RCM-PPE runs), though we suspect most groups would do this. We suspect that most EuroCordex RCMs will not have attempted to replicate the global model aerosol forcing in the way we have in the RCM-PPE runs (through EasyAerosol), because aerosol schemes in the driving and regional models will often be structurally different.	
75	4.2 The 12km regional climate model	In general this yields a poor description of lake temperatures, not only for the Swedish lakes but virtually all small lakes. Even most major European lakes have rather different winter temperatures and spring temperatures than the nearest sea-surface grid point.	Prof Dr Christoph Schär	We agree. In practice, our key consideration was a relative judgement: whether converting lake points to land would likely be better or worse than using SSTs from the nearest sea point. This was not well expressed in the original text. The revised version reads: " Therefore,the RCM land-sea mask was edited to keep only lakes for which the nearest GC3.05 sea point was expected to give a better representation of surface temperature than would be achieved by excluding the lake point."	
76	4.2 The 12km regional climate model	How was soil moisture initialized in the RCM experiments? This is critical as there does not appear to be a spin-up period.	Prof Dr Christoph Schär	The soil moisture was initialised using data from an ancillary dataset of observed soil properties (details in Walters et al., 2017). There was a 2 year spin-up period prior to 1980, which was discarded. This should be enough to spin up soil moisture, at least in the root zone (e.g. Guillod et al., 2017). The spin-up was already discussed in the experimental design section (4.3). We've revised the text to mention the initialisation.	
77	4.4 Evaluation of regional simulations for 1981-00	Fig. 4.3b: You need to specify whether these are all-day or wet-day percentiles, i.e. whether percentiles re taken over all days or only days above some precipitation threshold. In addition, I think that an intercomparison of precipitation frequency (P>threshold) would be desirable, as this is one of the standard metrics for evaluating precipitation climate.	Prof Dr Christoph Schär	Fig. 4.3b shows all-day percentiles. This has now been stated explicitly in the text and Figure caption. The RCM-PPE runs show a positive bias in wet day frequency, in common with most EuroCordex runs, and most climate models in general. We have added a paragraph after the discussion of Fig. 4.3b, quoting UK averages for obervations and the RCM-PPE ensemble-mean. We decided not to show maps (to keep length down), but we point out that the positive biases apply almost everywhere in the UK.	
78	5.1 Comparison of projections from Strands 1-3	Major comment: The normalization used in Fig.5.1 figure is useful as it isolates the differences due to regional variations. Yet I am very surprised to see that a corresponding figure without normalization is missing. It is evident that differences between the modeling approaches are much larger than it appears from the current display. The recognition that normalized changes are rather similar is not new but already discussed in AR5 (Chapter 12, see Fig.12.10 and discussion).	Prof Dr Christoph Schär	Fig. 5.1 is there to provide basic information about patterns of change, complementing the global mean information in Fig. 3.19. As noted in response to comment #240, we already provide comparisons of full (non-normalised) changes in Fig 5.2 (as spatial maps for a given period) and Figs. 5.3 and 5.4 (as time series). So we think this point is already covered, in ways that convey more information than we would get by simply repreating Fig. 5.1 to show national-average changes for some given period. We have acknowledged AR5 Chapter 12 (Fig. 12.10) as suggested, by citing it in the discussion of Fig. 3.21, which is where normalised changes first appear in our report.	
79	5.1 Comparison of projections from Strands 1-3	The numbers on this line appear to belong to the 90th percentiles of projected changes. Please also quote figures for median changes.	Prof Dr Christoph Schär	Done.	
80	5.1 Comparison of projections from Strands 1-3	Precip reduction of -27.5% (Strand 3). How does this compare against Fig.B1, bottom panel, where the median for Southern England is found at -50%/	Prof Dr Christoph Schär	Fig. B1 gives a Strand 1 example for 2081-2100, whereas Fig. 5.2d and associated numbers are for 2061-80, so they are not directly comparable. However, Strand 1 and 3 changes for a common period are shown in Fig. 5.2d, which the text goes on to discuss. There, we focus on comparing Strand 3 with 1 and 2 at the 90th percentile, as that seems the most apposite way of pointing out the consequences of not including CMIP5-based information in Strand 3. The median responses in Strand 3 are also generally drier than Strand 1 (esp over Enlgand & Wales), though the contrast is smaller cf the 90th percentile. We've added a sentence to this effect.	
81	5.1 Comparison of projections from Strands 1-3	Major comment: It is important that an effort is undertaken to compare the precipitation projections against the CORDEX simulations. Actually it appears that UKCP18 projects a significantly stronger JJA drying than CORDEX. According to page 79 line 29, Strand 3 projects a median drying of -27.5%. For CORDEX, only one out of the 15 analyzed 12 km models projected such a large UK summer drying, with a median change (over the Britisch Isles) amounting to merely about –14% (see Rajczak and Schar. 2017, Fig. S16). It is evident that the analysis of the CORDEX simulations gives a different projected range. It is also evident from Fig.5.3a and Fig.5.5a that some of these differences are due to the systematic differences between GC3.05 and CMIP5 ensembles. Ref: Rajczak J. and C. Schär, 2017: Projections of future precipitation extremes over Europe: a multi-model assessment of climate simulations. J. Geophys. Res. Atmos., 122 (20), 10773-10800, http://dx.doi.org/10.1002/2017JD027176	Prof Dr Christoph Schär	The discussion of Fig. 4.5d in section 4.5 has been extended to include a discussion of how the Strand 3 results compare against EuroCordex. This is also referred to in section 5.1, in the discussion of Fig. 5.2d. Indeed the EuroCordex ensemble (we focus on the 12km one, which is driven by five CMIP5 models, including for members of CMIP5-13) covers a spread ranging from small increases, to drying at about the median level of Strand 3. In winter, the correspondence is much closer - this has also been noted).	

83	3 Strand 2: A new ensemble of global climate realisations from 1990-2100	This part of the Land Report is extremely well written, and exhaustive. I read the whole thing without finding anything significant to comment on, UNTIL I started subsection 3.4, which presents the results of the ensembles evaluation. My comment here is very general, I'm not sure anything can be done about it, but I'm very concerned. I'm concerned about the behavior of the GC3.05 model, in its STD version and perturbed versions, for the historical period. The cooling for the most part of the 20th century at odds with observations, and the large regional biases connected to it, present a great challenge for the communication of these products, I believe. I'm wondering if, given that this behavior is the same in the STD version, single forcing experiments could be used to support the explanation given (aerosols) for this behavior. I'm not suggesting to run additional experiments for the UKCP18 project. Certainly single forcing experiments must be underway with the model in question. Having studies written up to which the authors of the UKCP18 project can refer, to prove what seems at this stage speculative, at least in the way it is described in this document, could help support the argument. I don't think that would solve the problem, but it could help. I'm also aware that the behavior here described is common to other models with interactive aerosol treatment. So the community has to reckon with it, not only the UKCP18 project. But the use of these results for practical applications presents a special challenge, and, stating the obvious, I want to underline the need for careful communication/justifications around this issue. Also the fact that the GC3.05 PPE has such a narrow spread, and a large portion of it (60%) comes from the perturbation of the concentration trajectories rather than the physics parameters perturbation is of great concern to me. I cannot evaluate if this latter result should suggest a failure of the PPE approach, or if the nature of the model is so rigid not to allow larger spread on the basis of the parameters chosen, and their values. But the small spread, together with the non overlapping quality of the two ensembles (GC3.05 and CMIP5) and their position with respect to the Strand1 envelope (Figure 3.19) will make the whole think difficult to digest by the users, I would think. I hope the communication of these products will be up to the challenge.	Dr Claudia Tebaldi	Thanks for the positive comments. We agree about the challenges, and that more experiments should be run. We would obviously have preferred to see a broader spread from GC3.05-PPE, though we would point out that there is more spread (relative to CMIP5-13) in many regions, compared to the situation for GMST. We do point out (and have emphasised this more in the revised version) that Strand 2 is not built to support statements of probability (see #50), and that Strand 1 can be used as context for assessments of impacts, or storylines, derived from Strand 2. Sections 5.1 and 5.2 contain discussion of ideas for how to interpret and use the Strands as a combined set of products. but we agree that an ongoing communication effort will be important. Regarding future experiments, several are possible, and we would welcome the panel's views on priorities. Ideas include control simulations to estimate drifts and provide a baseline for idealised forcing experiments to determine ECS and TCR values. Single-forcing experiments (either in coupled mode or as "Hansen" atmosphere-only experiments to determine forcing) would certainly be of interest too. However, some prioritisation will likely be needed.	
90	3 Strand 2: A new ensemble of global climate realisations from 1990-2100	"Potentially important" is an optimistic way to portray a situation that is fundamentally a challenge in interpretation. This will need very careful handling through communication/guidance/caveats.	Dr Claudia Tebaldi	We agree. Here we are only raising the idea that Strand 2 can be considered for use in international assessments, rather than making this a major focus of the report. However, we have added the sentence : "In such applications, further evaluation of how well the individual Strand 2 realisations simulate regional drivers of changes in temperature, precipitation and other key variables will be an important precursor to using the results for impacts assessments."	
95	General	The document seems to be more of a sales brochure than impartial information. The caveats and limitations document provides some information on the conditionalities and disagreements but this document bigs-up the comprehensive nature of the work. This leads to confusion; is the information reliable and brilliant (as the science document very strongly implies) or is it contentious, debatable and requires cautious application (as the caveats and limitations suggests). An approach to the documentation which was more reflective of the caveats and uncertainties in the approach would be more informative and helpful to the user when considering how to use the information. As it is, the document easily falls into a format which is written like a sales document. It seems to be trying to sell ukcp18 rather than informing about its pros and cons in a balanced and impartial way. Maybe that is inevitable because the same organisation which produced the projections is writing the document i.e. maybe the Met Office should "sell" ukcp18 it is the result of all their hard work. In that case, however, others should be asked to write impartial critiques that are provided by DEFRA alongside Hadley Centre Documentation. Whatever the case this sales aspect is a source of confusion and misdirection. I shall refer to this aspect as "sell" or "sales".	Dr David Stainforth	We agree, of course, that the document should provide a balanced assessment of the strengths and limitations of the three Strands. It certainly shouldn't be a sales brochure, and we have not written it with that intention. Indeed, overall we don't agree with David's view that it reads like one. Many examples of model errors are provided and discussed, for example, and limitations and conditionalities of the Strands are also discussed, particularly in sections 5.2 and 6. Most of David's specific comments relate to concerns around the probabilistic projections. In the original draft we did (in an attempt to keep the size of the report down) focus mainly on how the method for that product has evolved since UKCP09, reiterating only briefly the material already published in UKCP09 and the specialist literature. This had been done with the idea of reserving more space for Strands 2 and 3, which are both completely new. However, we agree that a recap of the issues David raises should be provided earlier in the document, to explain what we think are the main benefits and limitations of the pdfs before readers encounter the results. So we have included a discussion in section 2.1, under the heading "interpretation of the probabilistic projections." In this, we discuss the main conditionalities and assumptions in the method. We explain that we see the main benefit as a broader exploration of uncertainty than can be obtained from CMIP ensembles in isolation (for which there is plenty of specific evidence), and we discuss the issue of how we treat structural uncertainties (by using CMIP5 models as proxies for the real world, and stating that this neglects by construction the unknown effects of common systematic errors). We explain our rationale for this, and quote the Frigg et al (2015) paper (on which David is a co-author) to note their disagreement with this approach. We also note that the presence of common errors in models is an issue for all projections, not just probabilistic ones, and that a judgement that common errors are not large enough to question the usefulness of climate model projections is a widespread one. We agree with David that unqualified adjectives like "robust" or "comprehensive" should be avoided - these have been removed. We have also removed "relatively comprehensive" (which reflected the point about broader-than-CMIP ranges made above), since short-hand phrases like this may be unclear without repreated re-statement of precisely what is meant.	
96	General	A further, related, general point is that the document implies a number of conclusions without necessarily stating them directly. I think it is important to evaluate the document on the message that a reader is likely to take from it, not just on the factual accuracy of the particular phrasing used. As such I caution against accepting responses to any of the reviewer comments which are founded simply on factual accuracy rather than what is likely to be understood by the reader. Points such as (5) and (8) below fall into this category. The text referred to in these comments gives an impression that ukcp18 is comprehensive and supports risk assessment while actually only stating that it was designed to do these things. Such statements should be assessed on the impression given.	Dr David Stainforth	As explained in #95, we have revised the text to remove phrases such as "relatively comprehensive", and have explained more explicitly that we see the benefits of Strand 1 as provididing a broader view of uncertainties than is available from CMIPx ensembles alone. There is ample evidence of the latter, both in this report and previous literature. Having said that, we think that it is important to stick to the discipline of factual accuracy in our statements of purpose (hence the "designed to" wording in some places in the report), precisely because we do *not* wish to run ahead of the evidence and claim successful delivery before we have presented supporting evidence. Therefore, we don't think this comment is a fair criticism, and we believe that we should trust readers to understand the aims, assess the evidence we present, and then form their own judgements regarding fitness for purpose, rather than jumping to conclusions in the way David suggests.	

97	1 Introduction	"as comprehensively as possible" This is the start of the sell. This gives the impression that this approach is the best that one could possibly imagine. That is one viewpoint - possibly the viewpoint of those who produced ukcp09 - but it does not represent some sort of scientific consensus view. As such it should not be read in the same light as the scientific consensus over the reality of anthropogenic climate change. The claim is actually that it was “designed to represent uncertainties as comprehensively as possible” but the point still stands. The claim that it was designed to do this will be read by most as an indication that it actually does do this. In fact some other scientists take the view that the design does not represent uncertainties as comprehensively as possible. Surely this point is important to communicate to the users. By P4L12 the implication has become that uncertainties are explored comprehensively, not even “as comprehensively as possible”. This is unjustified sell.	Dr David Stainforth	We have replaced "as comprehensively as possible" with the more specific phrase "consistent with available climate model simulations and the knowledge contained within them". This is the same wording we use at various points, later in the text. The "designed to" wording is there simply to avoid the impression that we are telling readers that we succeeded, whereas in this text we are simply saying what the purpose was. Any judgement of success or failure should be made by readers and users - if they find the pdfs credible and useful, based on th evidence presented, they should use them. If they don't, then they shouldn't.	
98	1 Introduction	The use of phrases like “designed to support risk assessment” give an impression of reliability while not strictly claiming it. They are likely to give users a misrepresentative impression of the situation. In this case, ukcp09 was indeed “designed to support risk assessment” but the extent to which it can be used to that end is a source of substantial disagreement. Statements like this ensure that the existence of such debates and disagreements are unhelpfully avoided. The impression is given that not only were they “designed to support risk assessment” but that they can support risk assessment. I would argue with such a claim but the important point for the documentation is that it would be better for UK adaptation to climate change if such arguments were to take place in the open rather than for them to be dismissed through the argument that the documentation only claims that the product was designed to support risk assessment. This aspect of semantics has been a problem in the past. The consequence is that this sort of statement is the worst of all worlds. The claim that ukcp09 supports risk assessment would need to be accompanied by caveats that such a claim is subject to disagreement but that would be much better than the current phrasing which obscures the state of scientific discussion/debate.	Dr David Stainforth	We use phrases such as "designed to support ..." in places where we are describing the *motivation* for the various components of UKCP18 (not just the probabilistic projections). This seems appropriate to us. Such statements of motivation are followed by a set of evaluation material, which readers can use to decide whether they find the evidence strong enough to use the Strands for the intended purposes. David seems to be assuming that readers will jump straight from the introductory discussion to a conclusion of fitness for purpose, without reading the intervening evidence. We don't agree with him. We would also emphasise that we are not suggesting that the UKCP18 results are the *only* evidence that stakeholders can or should use. For example, many stakeholders used the UKCP09 pdfs in the first and second CCRA assessments (thus, presumably, not sharing David's misgivings), however they were not limited to exclusive use of the pdfs. Indeed, CCRA2, in particular, featured use of a broader set of information and approaches. This is why we use the word "support", rather than stronger terminology which might (wrongly) appear to rule out consideration of other lines of evidence. We have included more discussion of the latter point, in response to # 149.	
99	1 Introduction	Sell. The probabilistic projections are claimed to be “robust”. I disagree. The caveats and limitations document raises some of the reasons while also obscuring some of the reasons. Even as they stand, however, they are sufficient to undermine a claim of “robust”.	Dr David Stainforth	Deleted robust. We agree that this should be qualified and expanded upon, but in this scoping section it is too early in the document to provide the discussion needed, so simpler to drop it. See response to #99 above.	
101	1.3 Three strands of UKCP18 land projections	Sell. “comprehensively”. See point 3 above. The implication is that this is a comprehensive assessment of uncertainties. This goes beyond the claim for ukcp09 that it was as comprehensive as possible. I don’t think it is justified.	Dr David Stainforth	Deleted comprehensively. As with #101 above, we agree that this needs qualification, but again, the length of discussion needed to do so would be out of place here.	
102	1.3 Three strands of UKCP18 land projections	The claim that this supports future risk assessments implies reliable or reasonably reliable probabilities in strand 1. I do not believe that to be the case. The caveats and limitations touch on the possibility of this not being the case while here such important caveats are simply ignored. This conflict in the documentation leads to a situation in which we can not expect users to understand how to use the information.	Dr David Stainforth	We've reworded the sentence to make the link more explicit between the first part of the sentence "that the probabilistic results provide support for future risk assessments" and the second (revised) part "... by providing context for other sources of information that represent uncertainties in a more limited way." This reflects, for example, the extensive use made in the CCRA1 and 2 risk assessments of the 10-90% probability ranges (rather than more detailed assessments of relative probability within the range). Our point is that these ranges are generally broader than ranges derived from other climate modelling products, such as CMIP ensembles of opportunity. Therefore, the pdfs are useful in reducing the risk of overconfident interpretations that might occur in their absence, if users only had small ensembles of multi-model or PPE results to rely on. This is supported by UKCP09 and post-UKCP09 literature, as well as by the new results in UKCP18 itself. So, with this clarification, we don't think the statement runs a risk of implying that an excessive level of precision should be attached to the detailed varaitions in relative probability as we move through the Strand 1 distributions for some chosen variable. The latter issue is picked up below.	
103	1.3 Three strands of UKCP18 land projections	This statement is an example of confusion in the process of climate change prediction. The impression it gives is that as a result of the extra complexity the information is better and more useful. This is true only if the range of future outcomes conditioned on no carbon cycle feedbacks are themselves reasonably reliable. Some people do not believe that to be the case. If this is not the case then adding ESM information, while providing a more complete view, does not provide better information since it suffers just as greatly from methodological failings. This, then, is an example of further sell. The communication that more processes are considered of course leads to the impression that the results are more reliable but that is probably not the case.	Dr David Stainforth	All climate projections (even those based only on physical model) implicitly include some estimate of carbon cycle feedback, because this is required to convert carbon emission pathways into atmospheric CO2 concentration profiles. In UKCP09 we included *uncertainties* in carbon cycle feedbacks in the probabilistic results. In doing so, we were following the judgement of IPCC, which had assessed the C4MIP ensemble (the first coordinated multi-model ensemble of carbon cycle projections), and had provided an assessment of uncertainties based on its results. CMIP5 followed suit by including a second generation of earth system model projections, which were assessed in AR5. We also, therefore, follow suit in UKCP18. Our approach is that process uncertainties should be included as an intrinsic component of the projections in cases where capability has been developed to the point at which: (a) multiple modelling groups are able to produce alternative simulations, and (b) confidence in the results has developed to the point at which the international community feels able (through papers, and coordinated consensus assessments such as IPCC) to provide assessments of uncertainty and impact. While the results are of course conditional on the chosen methodology (a point we make in the text), our judgement is that to exclude carbon cycle ensemble evidence from the data would exclude a major *known* source of uncertainty, thus making the results knowingly overconfident by construction, and therefore inconsistent with the understanding included in the available climate modelling evidence. We don't think this would be a defensible position. In contrast,we believe that methodological dependencies, or neglect of uncertainties that fall into the category of "known unknowns" (such as unquantifiable effects of common systematic biases) *are* defensible. They should be clearly stated, but they are inevitable, and common to all national scenarios produced around the world.	
104	1.3 Three strands of UKCP18 land projections	Sell. The claim that strand 2 is particularly suited to the development of storyline approaches is a distortion of the storyline approach as presented by Hazeleger et al. in Nature Climate Change which proposed the idea of developing physical justifications of potential changes and filling in the details with models. A key aspect was that one did not start by being constrained by what models (even PPEs) can and can’t simulate. It is very disappointing that ukcp18 seems to have adopted only a very limited and constrained idea of the storyline approach.	Dr David Stainforth	This text was aimed more at the development of physical narratives in the sense of understanding large-scale drivers of regional changes (Zappa and Shepherd, 2017), than the idea of carrying out shorter simulations to look at how specific types of event might change, in a framing of climate change derived either from models or theory (Hazeleger et al). We have amended the text accordingly. We do refer to the tales of weather idea more explicitly in section 1.2, so we don't think it's fair to accuse us of distorting their discussion. We also support the point that a range of approaches to decision-making can and should be considered, including vulnerability-led or narrative approaches, as well as "predict-and-provide" from model projections. However, the purpose of the science report is to describe the model predictions and how they can support stakeholder activities, whereas wider questions around alternative decision processes are, we think, in the realm of user guidance.	

105	1.3 Three strands of UKCP18 land projections	Sell. The text highlights that Kendon et al., 2017, suggest that one should be cautious in interpreting their results. It goes on to throw such caution to the wind. I don't understand how an adaptation planner could usefully interpret such information. The text implies that the caution suggested is not something to be concerned about while the fact that the issue is raised suggests that this aspect of ukcp18 is not fit for purpose. There is a strong sense of having ones cake and eating it here. And indeed throughout this document.	Dr David Stainforth	The 2.2km results are not presented in this report. They are mentioned only to provide users with advance notice of them. Of the two cautions we mention, one relates to the need to move beyond individual "this might happen" simulations, to production of ensembles that provide some information on uncertainties. This is a basic requirement for use in impacts assessments, and will be addressed in UKCP18. The second caution is a resolution-related caveat relating to not capturing smaller showers, that we think users will need to be clear about. It means of course that the simulations (like any climate model simulations) won't be perfect, and we've revised the last sentence of the relevant para to reflect that the degree of utility will be dependent on the skill of the runs. However, we disagree that discussing what the simulations might be used for is somehow tantamount to ignoring the cautions cited above. Rather, we think this is typical of how climate model ensembles are presented for potential understanding and use: There is likely to be evidence of skill, mixed in with evidence of some systematic biases, and users will need to assess the implications of those on a case-by-case basis. So we think the message needs to be "use with appropriate caution", but not "wait till all caveats have been resolved".	
106	1.4 Emissions scenarios	I have not yet finished reading the caveats and limitations but I hope there is a reasonable discussion of the debates about the relevance and usefulness of pattern scaling methods such as the paper by Lopez et al. in Climatic Change 2013. There are certainly significant doubts and concerns over this issue which should be raised in a document like this if it is to provide relevant information to the users.	Dr David Stainforth	In this report, limitations of pattern scaling are discussed in section 3.5. Its application in the synthetic projections is discussed in the Good et al. report.	
107	2 Strand 1: Probabilistic projections	A key point to communicate is that there is scientific disagreement over this methodology and whether it can/should be used in risk assessment. The caveats and limitations document is important but it should be a document which gathers these issues together so they can be understood as a whole. It should not, however, be a replacement for raising these issues in the rest of the documentation where they matter for understanding whether the method and results are robust and accepted within the scientific community.	Dr David Stainforth	See #95. The new "interpretation of the probabilistic projections" text in section 2.1 discusses the Frigg et al. paper, and the wider assumptions behind use of climate model projections to support stakeholder activities. We would note, however, that the Frigg et al. paper is purely an opinion piece. It repeats at length material from previous papers describing the UKCP09 methodology, to which the authors add their views (unsupported by any specific scientific evidence), questioning a number of assumptions in the method. We note that the UKCP09 methodology papers have been available since 2012 and 2013. In that time, no-one has, to our knowledge, published a journal paper containing hard, reviewable evidence in support of any critique of the methodology. We should, and do, provide general discussion of the assumptions behind the method, and conditionalities and potential concerns that follow from them. However, we believe that the main discussion of strengths and limitations should be evidence-based. There are a large number of published examples in which climate scientists have decided to use the UKCP09 pdfs, or pdfs for other parts of the world derived using the same approach (such as in the FP6 ENSEMBLES project). There are no published examples that we know of, in which the pdfs are demonstrated to be misleading, or not useful, in the types of application for which we would expect them to be useful. Whilst we acknowledge David's views, and now cite the Frigg paper in the text, we do not think it would be appropriate to include text suggesting that opinion around the usefulness of the pdfs is somehow split 50/50, because we do not see published evidence of such a position.	Following discussion at the Moderation Meeting, we added some general comments on the nature of probabilistic climate projections at the start of section 2.1, and shortened the extensive "interpretation" discussion that had been added in the main body of section 2.1. However, we retain a discussion of the main conditionalities and assumptions, and cite Frigg et al as an example of how these are subjective, and open to critique.
108	2.1 Overview	I contend that integrating over parameter space in this way leads to largely meaningless probabilities as far as the real world is concerned. I also contend that relative likelihoods are misleading unless accompanied by some indication that the absolute likelihoods are not very small. These are some of the fundamental methodological flaws in strand 1. There are papers discussing these issues. I'd be happy to discuss them with DEFRA.	Dr David Stainforth	Sexton and Murphy (2012) examined the sensitivity of the pdfs to a range of alternative assumptions, including changing the expert prior distributions for model parameters, increasing the structural uncertainty, and changing the constraining observables. These results changed details of the pdfs, but had only modest impacts on their main properties, such as 5-95% probability ranges, or the result that outcomes near the median always attracted higher relative probability than those near the extremes. Furthermore, Sexton et al. (2012) demonstrated that the recent climatology observables used to constrain the UKCP09 projections lay within the spread of modelled outcomes. This is also the case for UKCP18 (more detail will be given in the specialist journal paper to follow, Harris et al (2019)). This is also the case for the historical change observables, such as GMST (see Fig. 3.19) and CO2 concentration (Booth et al., 2017). Therefore, we disagree that the relative likelihoods calculated in the method are misleading, because we have provided published evidence to the contrary. It does, of course, need to be acknowledged that there is an enormous amount of observable data available to evaluate climate models, and of course examples can be found of common biases. The likelihood calculation of Strand 1 would indeed not be well constituted, if we were to focus it on some subset of poorly-modelled variables, essentially because a model variant ten standard deviations away from a poorly-modelled observable is not really any better than one that is eight standard deviations away. Our approach, therefore, is to use a broadly-based basket of observables that cover multiple aspects of climate at large scales. It turns out that models these simulate well enough to allow a tractable likelihood calculation, which is why we think the methodology is justified. Importantly, we then add the caveat that the effects of common errors are an unknown factor that lies outside current knowledge. We have included discussion of this point in the new text placed in section 2.1. Overall, we think that this is an approach consistent with the conditionalities that we state, and with the widely-held view in climate science that "all models are wrong, but some are more useful than others".	
109		Sell. I disagree that the uncertainty ranges are as comprehensive as possible and I think there are severe risks in using them to inform risk assessments. These are contentious statements. In any case it is misleading to claim that it is the emissions-driven approach that ensures this.	Dr David Stainforth	The "as comprehensively as possible" phrase was an attempt to express briefly the point that the pdfs should represent the known uncertainties that are currently represented in available model ensembles (which include carbon cycle processes, via ESMs) as fully as possible, coupled with the unavoidable neglect of known or unknown unknowns - processes yet to be included in models, or the effects of common biases, which cannot be quantified a priori because the earth system is too complex to allow this to be done by expert judgement alone. This broader rationale is now explained in text added below the bullet identified in this comment. We have reworded the relevant bullet with a longer sentence which expresses the point in relative terms, to avoid any risk of implying overconfidence: "Both types of feedback contribute substantially to uncertainties in projections of global mean surface temperature (GMST) (e.g. Knutti et al., 2008), which in turn influences the spread of projected regional changes. Accounting for carbon cycle uncertainties therefore ensures that the probabilistic projections take fuller account of known limitations in the current modelling of key earth system processes, and are therefore more useful in informing risk assessments."	

110	2.1 Overview	Sell. These lines embed the idea that relative likelihood is what is relevant. In practice one can have a highly competitive model which is still highly unlikely to represent the complexity of processes necessary to make climate projections about reality at local scales. The claim of competitiveness is a distraction which is clearly meant to give confidence to the users but in fact it should not be taken that way. The documentation should: a) communicate that there is strong disagreement on this issue, and b) provide information on relative likelihood c.f. absolute likelihood and the absolute degree of consistency of these models with historical reality. This should really be done across a range of variables and at a range of scales.	Dr David Stainforth	We disagree that this is a "sell". The statement here is addressing the question of whether HadCM3-based PPEs (being built from a CMIP3-generation model) remain reasonably competitive with the current state of the art. This question, and its importance, is raised earlier in the relevant paragraph, so we think the specific context will be clear to readers. Regarding the comments about disagreements and relative likelihoods, see #107 and #109.	
111	2.1 Overview	Later in the documentation it is highlighted that the strand 2 and 3 ensembles were not designed to explore uncertainty and shouldn't be interpreted probabilistically. The same is true for CMIP5. Consequently this discussion of smaller and larger chances is worse than meaningless because it contradicts the later discussion of how ensembles can be interpreted and presents a confusing message.	Dr David Stainforth	Reworded to avoid using the language of probability in describing the CMIP5 results: "An exception was that CMIP5, for England and Wales, provided fewer realisations showing substantial future reductions in summer rainfall (i.e. reductions of exceeding 10% per degree of global warming), and more simulations showing modest increases, than suggested by the corresponding UKCP09 probability distribution."	
112	2.1 Overview	This sentence somewhat encapsulates the contradictions within ukcp18. The document communicates a message of confusion. On the one hand it sells strand 1 as comprehensive and suitable for risk assessment while on the other it throws in suggestions that they should be used cautiously and verified against other information. If the information in strand 1 cannot be taken at face value then it should be explained up front why this is the case. It is insufficient to simply say that strand one results should be compared against alternative projections. Why? If strand one is reliable there is no need. If it is not then how unreliable is it; is there any reason to take any note of it at all? These are the issues that I have tackled in my own work. My own view is that the methodology behind strand 1 is sufficiently flawed that it could be very misleading for risk assessments. Presumably those responsible for producing ukcp18 take a different view but it is crucial that they are more explicit about the extent to which it can be relied upon. Statements such as this sentence throw the responsibility back on the user who can not possibly have sufficient knowledge of the methodological processes to make an informed judgement. The likely consequence is that a user either ignores these warnings and is persuaded by the "sell", or they note these warnings and ignore the details of strand 1. The former leads to dangerous over interpretation and the latter makes the high cost and complexity of the data analysis in strand 1 mostly pointless. It is perhaps the responsibility of DEFRA rather than the Met Office to grasp this nettle and ensure a wider and more inclusive multi-disciplinary debate.	Dr David Stainforth	We don't agree that this message is contradictory. Given that the pdfs are communicated as being conditional on their inputs (a point we made more than once in the draft text, and now more often in the revised version), then we don't see why a stakeholder shouldn't consider other lines of evidence, conditioned on different inputs, which could reveal possible outcomes outside the range of Strand 1. We think this is a healthy piece of advice, which reflects practice (used successfully, for example) in the recent Climate Change Risk Assessment, or in impacts papers such as Kay and Jones (2012). It encourages users not to place an exclusive reliance on the pdfs. As explained in other responses, we provide clear evidence that users will see a broader range of plausible future climate pathways, and therefore run fewer risks of making overconfident decisions, with Strand 1 included. If we were to remove Strand 1, it would, in our view, have the perverse effect of increasing such risks. Of course, use of the pdfs does require a level of common sense, and there is always a risk that a minority of applications could use them in an over-precise way (though we have included commentary to discourage this). However, if the majority do use the results sensibly, then we would argue that the pdfs provide a net benefit to the UK adaptation effort. Given that this industry now has nine years experience in using the UKCP09 pdfs, and have taken them up in many studies without (to our knowledge) any obvious evidence of widespread misuse, we think it is justifiable to assume that they will continue to provide a useful contribution. Nevertheless, we do agree that considering different sources of information requires care and insight (we discuss these issues in section 5.2). But we think that trust should be placed in the expert-practitioner community to understand and use the products appropriately, with support from the user guidance and other outreach activities.	
113	2.1 Overview	It's unclear at this stage how this allows the number of PPE inputs to be reduced from seven to three.	Dr David Stainforth	Added: "... mainly because ocean, aerosol and carbon cycle uncertainties can now be quantified from one ensemble, whereas separate ensembles were used in UKCP09 (Murphy et al., 2009; Harris et al., 2013)."	
114	2.1 Overview	I reiterate the need to provide information on absolute likelihoods rather than just relative likelihoods.	Dr David Stainforth	See response #108.	
115	2.1 Overview	This seems to be claiming that strand 1 produces probabilities of historic and future change (conditioned on emissions, of course). That indeed seems to be the case when one looks at the results. Can these probabilities be used as probabilities in risk assessments? P8 L21-22 implies they can. P9L16-18 suggests perhaps not. What exactly is being claimed for strand 1? There is a need for clarity.	Dr David Stainforth	See response #112. We don't think there is any contradiction between the two statements David points out, essentially because the pdfs are conditional (and are acknowledged as such). They are Bayesian statements of the degree of belief in some set of climate outcomes, dependent on the evidence put into the calculations. This is how we describe and assess them in this Science report. We also think, of course, that they are useful for risk assessment. But that doesn't mean that users should necessarily just plug the probabilities directly into (say) a cost-loss model in a decision process, without considering other evidence. This is somewhat akin to the process IPCC goes through, in distinguishing between likelihood and confidence statements. This is an issue that the user guidance might usefully discuss, but we think that usage in decision frameworks is outside the scope of the present report.	
116	2.1 Overview	There should be an explanation that this claim is a source of scientific contention.	Dr David Stainforth	See response to #95 above. The discrepancy term, our rationale for estimating it the way we do, and the main caveat relating to neglect of errors common to all models, is now discussed earlier, in an expanded section 2.1.	
123	2.2 Technical summary of methodology	Why should this "discrepancy" relate to the relevant one i.e. the distance between the models and reality rather than the PPE and another ensemble? Disagreement on this point should be highlighted.	Dr David Stainforth	See #95. We note also that Sexton et al (2012) provided evidence that the joint prior distribution of historical climate outcomes was found to compare adequately (i.e. encompass) the constraining observables. This provided evidence that the historical component of discrepancy (which has to be determined a priori, to avoid double-counting the observations themselves) was effective in broadening the space of outcomes to a sufficient degree. The future component of discrepancy is obviously impossible to verify in a similar way. It is important to understand that discrepancy is a user choice in the methodology. If we thought it was possible to quantify the effects of unknowns, we would have done so. By taking the approach we do (using other models as proxies for the real system), we can present a clear interpretation of the pdfs as evidence that combines PPE and multi-model inforamtion, with unknowns and common errors treated as a caveat. This, we believe, is also consistent with the general spirit in which climate model projections (regardless of whether they are probabilistic or not) are made available for use in adaptation studies.	
124	2.2 Technical summary of methodology	Information should be provided on likelihoods of any of these models being consistent with reality so that the users can have a context for making their own judgement about whether the discrepancy term is likely to be meaningful.	Dr David Stainforth	See #108.	
125	2.2 Technical summary of methodology	Figure 2.3 and 2.4b should both be for southern England. This is important so that readers can follow how data is used and how results are connected. I understand that for geographical fairness one might wish to use different regions in different plots but for the purpose of communication and understanding it would be far preferable to use the same region in both.	Dr David Stainforth	Although geographical fairness is a consideration, we actually chose regions for each plot that best illustrated the relevant aspect of the methodology. We take the point about connection, but we'd prefer to keep the use of different regions, to make sure that the impacts of different steps in the method are made clear.	

128	2.2 Technical summary of methodology	Why do the HADCM3 analogues overestimate carbon cycle feedbacks in this way?	Dr David Stainforth	There is a spread in the discrepancy estimates for GMST (light lue curves in Fig 2.4a, top right), and of course many factors (including physical feedbacks, time-dependencies in their strengths, as well as carbon cycle feedbacks and their components) could play a role in determining each of the estimates. We put in the carbon cycle quote to provide an illustration of one contributory factor, but a detailed analysis of case-by-case reasons would be a major undertaking. This is a very interesting question, but since our focus in this report is on illustrating methodology, we think the material provided is sufficient, and note this question as something to consider for a speclialist journal paper.	
129	2.2 Technical summary of methodology	By this stage in the document I found myself again asking who is the intended audience? It seems to provide substantially too little detail to provide scientific justification of the method or to facilitate constructive scientific debate. At the same time it seems to provide too little detail and assume too much specialist knowledge to provide meaningful guidance to even educated users. It doesn't seem clear what it is trying to achieve.	Dr David Stainforth	The target audience (see #227) is essentially scientific peers (essentially the IPCC Working Group 1 community) and expert users (essentially IPCC Working Group 2 scientists interested in impacts). As noted in #229, the basis of the Strand 1 methodology (which has not been altered by the updates for UKCP18) has been published for some time (Sexton et al., 2012; Harris et al., 2013), and the research community has had ample opportunity to scrutinise it. So we don't think it is at all fair to suggest that it is under-described. Our approach is to refer readers to the past papers for the fundamentals, and to focus the main detail provided in this report on the updates. We have revised section 2.1 to make this rationale clearer, and we have reprised some of the previous information in places, to aid understanding (see #209 and #211). This approach does, undeniably, place an onus on readers wishing to critique the method in detail to read the past papers as well as the current report. But we think this is unavoidable (given that combination of such a large amount of information in a sophisticated statistical fremework is inevitably a substantial task), and we believe that we have provided enough information and links to allow interested readers to review the method in considerable detail, if they wish to do so.	
130	2.2 Technical summary of methodology	similar thoughts	Dr David Stainforth	See #212. We have expanded the discussion of observational constraints, and promoted most of Appendix B to the main text.	
131	General	I have written before about the flaws in the methodology used in strand 1. I find it deeply worrying that there is no reflexion on the reliability of the approach taken nor mention that it is an approach which is far from representative of scientific consensus. Some of these issues are dealt with in the caveats and limitations document but that document should gather such caveats together, it should not absolve the authors from dealing with the particular issues as they become relevant (I have a series of comments on that document too, which I'll send later.)	Dr David Stainforth	See response to #95. We have added a broader discussion of the interpretation of the pdfs, and the assumptions, caveats and conditionalities behind them, at an earlier point in the report (section 2.1), to add to the discussion we had in sections 5.2 and 6. We have cited Frigg et al, though we note in this reply that this paper is an opinion piece, and does not, in our view provide any specific evidence that the Strand 1 method is flawed. We believe it is set up, and described, in a way that is consistent with the methodology and assumptions that we use.	
133	2.2 Technical summary of methodology	This is a key assumption which is dropped in as if there is no reason to question it. It is characteristic of the document and the methodology to make such assumptions without highlighting their questionability and potential impact.	Dr David Stainforth	Revised text to read: "The residual term (which accounts for interannual variability, as explained in Stage 2) is noisier than the bias term, hence it is difficult to justify expressing it as a function of GMST response. Therefore, it is simply assumed to be independent of scenario, and sampled as in the RCP8.5 case." It is unavoidable to use a simple assumption (as earlier text explains), since ESPPE simualtions are not available for RCP4.5 or 6.0.	
134	2.2 Technical summary of methodology	The "hiatus" of a few years ago demonstrated a severe miscommunication between climate scientists and journalists / the public over whether climate change meant monotonically increasing GMST. The kind of terminology used here, the emphasis on the change in the median, contributes to such miscommunication.	Dr David Stainforth	The use of 30-year means in this figure is motivated by its purpose, which is to compare against UKCP09. Hindcasts and projections of GMST at the annual time scale are considered in section 3.4 and 3.5, where the hiatus is discussed. However, we have added a footnote to avoid misunderstanding. This says " Note that at the annual time scale (e.g. Sexton and Harris (2015), Fig. 1), time series of individual realisations that comprise the probabilistic projections can show periods of 10-20 years during which GMST departs from a monotonic increase, due to the influence of internal variability. This is also true of the Strand 2 simulations (see Fig. 3.19)."	
135	2.3 Examples of the probabilistic projections	I might have the details wrong but I thought these simulations used RCPs from 2005. The suggestion that responses emerge in the 2030s is presumably related to a period after which the emissions began to differ which in this case is 2005. (How different are RCP2.6 and RCP8.5 between 2005 and 2018?) I suspect some users would interpret this as differences emerging in the 2030s between scenarios in which there is substantial emissions reductions from now c.f. little emissions reduction from now. In fact that contrast might lead to differences emerging a decade later. I'm just concerned that this doesn't communicate effectively. I'm concerned that this will lead to misunderstanding of when different policy actions can lead to differences in climate.	Dr David Stainforth	We changed "emerge from the 2030s" to "emerge from about 2030 onwards". We also aded a footnote to explain the time scale. This reads: "Although the CO2 emissions pathway in RCP2.6 differs from that of RCP8.5 after 2005, emissions in RCP2.6 are assumed not to start reducing until after 2020, and CO2 concentrations in the atmosphere stay close to those of RCP8.5 until the mid-2020s. Hence, there is a delay in the emergence of clear differences in GMST response between the two scenarios."	
136	2.3 Examples of the probabilistic projections	Sell. This claim is entirely conditional on the methodology being presented. It could change completely (not marginally) with a different methodology. It is a flaw in the document that the conditionalities of the method are lost or ignored and instead we are left with conclusions which appear robust and will be carried forward as if they were so. This will lead to adaptation planning which is unaware of the conditionality of the information and thereby risks undermining the development of robust adaptation strategies.	Dr David Stainforth	We changed the relevant sentence to read: "In agreement with previous studies (e.g. Knutti et al., 2008; Hawkins and Sutton, 2009), the Strand 1 results indicate that uncertainty in emissions and uncertainty in climate response are comparable in their impacts on the range of plausible GMST outcomes during the coming century." This is a common and uncontroversial result, commonly accepted as current understanding in the climate science community. We think it is a gross exaggeration to suggest that the result could change completely, under a different approach. In our view, this could only happen if such a study was poorly constituted (e.g. by failing to account for model uncertainties in a reasonable way).	
138	4.5 Projected changes in the Strand 3 ensemble	In risk assessments one would surely want to use strand one as probabilities. This section makes clear that the UKCP18 providers think strand 1 can be used this way in risk assessments. The problem here is one of lack of clarity; a confusing message. Sections like this highlight the potential to use strand one as probabilities while elsewhere the document backs off from such a claim. It gives with one hand and takes with the other. If the producers believe these probabilities should be used as such then they should clearly and absolutely state this upfront. I believe this interpretation is severely misguided and likely to undermine UK adaptation initiatives. Nevertheless, clarity would help the users. It must also, however, make clear that this is a contentious view with scientific opinion divided. It should reference the Frigg et al., papers. What it should absolutely not do is make the misleading suggestion that the disagreement of interpretation is in regard to the fine details of the probabilities (e.g. is it 90% or 93%) but rather to the overall shape and location of the distributions (e.g. is it 70% or 30%).	Dr David Stainforth	See #95, 107, 112, 115 and 131. We have revised the relevant section of text to communicate Strand 1 thus: "a. The Strand 1 results are derived from large samples of potential outcomes and are formally constrained by a set of observables, allowing them to be interpreted as probabilistic estimates conditional on the climate modelling inputs and expert assumptions used in their construction (see section 2.1, also discussion in 5.2 below). " We are not aware of any alternative dataset, or evidence, that would suggest that the Strand 1 10-90 ranges should really be interpreted as, say, 30-70. Rather, we think the usefulness of Strand 1 is the converse: It reveals the limitations in sampling of uncertainties in other datasets, on which users would have to rely exclusively, if Strand 1 did not exist.	

142	5.1 Comparison of projections from Strands 1-3	Sell. This sounds like a big claim (“the most extensive range of information”), but what does it mean? The fact that it is the most extensive range of information doesn’t communicate very much apart from “trust us, this is great, don’t worry about the caveats”. In what sense is extensive good? Concise, reliable, easily applicable need to be achieved before extensive adds value. Government guidance should as far as possible be providing a dispassionate, objective presentation of the information in the context of wider scientific opinion. This document doesn’t achieve this.	Dr David Stainforth	We disagree. This is an evidence-based statement, not a "claim". UKCIP02 relied on single-shot scenario runs from only one climate model; UKCP09 had probabilistic projections and a set of regional climate model runs; UKCP18 has probabilistic projections, new regional model runs plus a new set of global simulations (Strand 2). The forthcoming convective-scale simulations will add further to the information base. Together, these lines of evidence provide a broader basis for applications work than either UKCIP02 or UKCP09. We also discuss conditionalities, biases and limitations in many places in the report, so we don't think "trust us, don't worry about the caveats" is at all a fair reflection of the balance of our discussion. We also encourage users to consider other sources of information (in section 5.2), where these might usefully augment the UKCP18 advice.	
143	5.1 Comparison of projections from Strands 1-3	I disagree – this evidence is far from being as robust as possible.	Dr David Stainforth	We have reworded this sentence to describe the Strand 1 results in the same terms as #97 above: "The primary purpose of the probabilistic projections is to ensure that UKCP18 includes a product focused on expressing known uncertainties in future changes, in order to reduce risks of overconfident decision-making in user applications. For example, extensive use of the UKCP09 probabilistic projections was made in CCRA2 (Humphrey and Murphy, 2016). Strand 1 is designed to provide ranges of outcomes consistent with, and limited by, the knowledge incorporated in existing ensembles of climate model projections (see section 2.1)." Whilst we agree that phrases like "as robust as possible" should be avoided, we don't agree with the implication in David's comment that something more robust could be done right now. In practical terms (of what data and knowledge is available to turn into specific advice at the present time), we are not aware that a *more* robust method of expressing uncertainties in UK climate has been published since UKCP09 came out, or that one is imminent from a different research group or project.	
145	5.1 Comparison of projections from Strands 1-3	This misses a central point of the Hazeleger et al. 2015 paper. See point 11 above.	Dr David Stainforth	Reworded to read: "•Provide a dataset useful for the development of narratives ...". As noted in #104, we are not suggesting that storyline-building should be based entirely on climate model output, merely that the Strand 2 and 3 data can contribute usefully. Again, we think the wider question of how climate model output is used alongside other approaches is mainly a question for user guidance and outreach, whereas here our task is to describe the model projections as a valuable but not exclusive source of information.	
146	5.1 Comparison of projections from Strands 1-3	This paragraph highlights the fundamental conflict in this document. On the one hand there are sales claims that these results are a core source of information for risk assessments accompanied by statements which seem to interpret them as probabilities. On the other hand there are suggestions that they can’t be used that way. This is a case of ukcp18 wanting to have its cake and eat it. How can a user be expected to interpret such information? Setting aside this fundamental issue I would make the following points. 1) These conditionalities should be upfront, not buried on P89. They are fundamental to any understanding of the results. 2) They should be highlighted in all results documents and web pages for the same reason. 3) The implication that this is the best that can be done with “current knowledge” is not justified. Certainly some people in the field believe there are better ways of using current knowledge. 4) It is very disappointing that the authors see fit to include this “disclaimer” deep in the document where it is likely to be missed by many. This reinforces my impression that the aim of the document is to sell the product rather than inform, guide and provide a context for these datasets.	Dr David Stainforth	This reflection of current knowledge point was also cited in section 6, but we have now added a longer discussion earlier, in section 2.1. See #95. As noted in #112, we don't agree that there is a conflict (fundamental or otherwise), between providing conditional pdfs demonstrating wide uncertainty ranges, advising users that we see these as an important tool for risk assessment (because they reduce risks of overconfidence), but also encouraging users to consider other information conditioned on different evidence.	
147	5.1 Comparison of projections from Strands 1-3	If the likelihoods aren’t precise how imprecise are they?	Dr David Stainforth	We have reworked this piece of text to clarify better what we mean. The point about precision is essentially that of Sexton and Murphy (2012), who showed the extent to which the shape and range of the pdfs depended on (sensible) variations to the expert assumptions on which they are based. This is somewhat distinct from the other point made in this part of the discussion, which is about considering other lines of evidence. We have revised the text to make this clearer. It now reads: "In UKCP09, Sexton and Murphy (2012) showed that the relative probabilities of different outcomes varied to a degree, in response to exploring plausible alternatives for expert choices such as the prior distributions for uncertain model parameters, the choice of observational constraints or the estimates of structural model uncertainties. None of these alternatives questioned the basic shapes of the distributions, and changes in the ranges of outcomes were modest. However, the results illustrate that Strand 1 users should consider the sensitivity of their application to reasonable variations in the probabilities provided, such as those outlined above."	
148	5.1 Comparison of projections from Strands 1-3	This seems to demonstrate a fundamental misunderstanding of statistics	Dr David Stainforth	We don't agree. However, we have reworded the relevant sentence to make it totally clear that we are suggesting that users consider the Strand 2 and 3 evidence alongside that of Strand 1, rather than that they should attempt to combine it statistically in some quantitative way.	
149	5.1 Comparison of projections from Strands 1-3	Surely it should be a core responsibility of the UKCP18 providers to provide guidance on this. How could users know more?	Dr David Stainforth	We have added the following text to make this point clearer: "This might include, for example, idealised studies of unlikely but plausible “climate surprises” arising from events or processes that are either not typically simulated by, or not yet included in, current climate models. Such events might include a future collapse of the AMOC, or a substantial release of carbon from permafrost, or of methane from ocean sediments (Collins et al., 2013)."	
150	5.1 Comparison of projections from Strands 1-3	How should a user decide what is a reasonable variation?	Dr David Stainforth	We think this point is better expressed by referring to other sources of information, rather than "reasonable variations". Text reworded to read: "More generally, the presence of systematic biases common to all climate models, and the incomplete nature of current understanding of earth system processes, underlines the importance of exploring the sensitivity of potential adaptation decisions or impact assessments to other sources of information beyond the scope of UKCP18." This sentence now followed by the new sentence in #149 above.	
151	5.1 Comparison of projections from Strands 1-3	I disagree with this statement. They are too restrictive.	Dr David Stainforth	Changed "particularly well suited to" to "a useful tool for". However, we stress again that our purpose in this report is to describe the model projections, and what they are useul for. This does not mean (and nowhere do we say) that development of storylines should be based *entirely* on model output. See #104. We think that broader questions around sourcing other approaches for supporting stakeholder decisions are an important issue, but one for wider stakeholder engagement activities, not the Science report on the model projections.	
152	5.1 Comparison of projections from Strands 1-3	Sell. Again the emphasis seems to be on promoting the results rather than informing about the difficulties and issues inherent in bias correction.	Dr David Stainforth	Added: "Note, however, that bias correction cannot remove the effects of large model errors on projected changes, and needs to be accompanied by a good understanding of relevant earth system phenomena, and how well they are represented in climate models (Maraun et al., 2017)."	

153	5.1 Comparison of projections from Strands 1-3	Already the conditionalities raised just 2 pages earlier are ignored.	Dr David Stainforth	Having stated the conditionalities earlier in this section, we are providing in this paragraph some advice on how the probabilistic projections might be used (in order to avoid overconfident inferences that might be drawn in their absence). We agree that the main underlying assumptions should be stated clearly. We think that stating the main caveats once up-front (added in section 2.1 in the revised version), and once again in the interpretation section (5.2) at the end, gives an appropriate balance that avoids both the risk of overselling that David raises, and also the impression of undue negativity that would occur if these were repeated every time the probabilistic projections are mentioned.	
154	5.2 Interpretation and use of the projections	The idea that such improvements are prerequisites implies an understanding of what level of model quality is required for this purpose. How has that been assessed?	Dr David Stainforth	We don't know of any way of assessing exactly how good is good enough, so we changed the text to read: "Such improvements in capability can help to justify use of such models to provide information useful for development of “storylines” ". We're just making the point that if a model simulates phenomena like (say) the Atlantic storm track better, then it is more likely that study of specific future storm events, and development of narratives of how they come about in a future climate, will be realistic and useful.	
157	6. Summary	This claim of reflecting current knowledge and modelling technologies gives the impression that the methods of ukcp18 encapsulate all relevant climate science and that they are acknowledged as such; updates will come but only because we will learn things that we don’t know today. I don’t accept that this is the case. It fails to communicate that the UKMO is the only organisation in the world pursuing this methodology and that other academics have serious misgivings with it; some being overtly critical. Current knowledge clearly leads to substantially differing opinions of the value and meaning of the information presented here. Updates are not a just matter of evolving capabilities but of fundamental debates over how to tackle these issues. The basis for these debates already exists. It is current knowledge. Furthermore there is severe danger in selling local climate projections as being uncontentious because it encourages the public to see them in the same light as the scientific basis for concern over climate change. This risks undermining wider communication of the risks of climate change.	Dr David Stainforth	Changed text to read: "Finally, we emphasise that the information in the UKCP18 projections reflects current modelling technologies and the knowledge incorporated in them." Of course, there is also theoretical knowledge, and some understanding of processes yet to be included in models etc, that forms a qualitative backdrop to the quantitative data provided in the projections. We note that all national scenarios use specific methodologies, so we think it is hardly surprising that the suite of UKCP18 products is unique to UKCP18. Specifically regarding Strand 1, it is true that no other centre produces probabilistic projections in a similar way (though some do include probabilistic estimates derived from CMIP ensembles). This does not necessarily mean, however, that the pdfs do not have an added value over reliance of CMIP results (as we argue in #95, for example), which other countries tend to rely on. See reviewer comment #51, for example.	
162	1.2 User and science drivers for UKCP18	Would be helpful to briefly describe the differences between GC3.05 and GC3.1 here, or link forward to section 3.2.	Dr Ed Hawkins	Forward references to section 3.2 (for GC3.05) and Appendix D (expanded to include a fuller discussion of how GC3.05 differs from GC3.1) have been included in section 1.2.	
163	1.5 Choice of baseline	Why not use 1981-2010, the current WMO standard baseline whch is used for many impact studies - this would enable UKCP18 to be put into a much broader context. I realise this would mean using a small portion of the RCP simulation, but the benefits would be worth considering. Also - some justification for the switch from a 30-year mean to a 20-year mean should be given.	Dr Ed Hawkins	Text added to clarify: "The standard baseline of the World Meteorological Organisation is currently 1981-2010. This was not adopted for UKCP18 because the RCP emissions scenarios used in the projections (section 1.4) start from 2006, hence the simulations include an element of predictive information during 2001-2010. However, users of data from Strands 1 and 2, and the 12km RCM simulations of Strand 3, will have the option to choose 1961-1990 or 1981-2010 as an alternative baseline." Another practical reason was the need to run the forthcoming 2.2km as 20-year time slices (including 1981-2000, with the option of filling in missing periods later), so we couldn't be sure of being able to provide 2.2km data for 2001-2010. At the time of writing, however, no decision has been made about how the 2.2km data will be launched, so we have avoided referring to that aspect in the text.	
164	3.3 Design of ensemble projections	How many of the final retained PPE members were in the 442 selected members, and how many from the added 115 members?	Dr Ed Hawkins	Thirteen came from the 442 surviving the TAMIP evaluation, plus two from the additional 115 AMIP members. This has been noted in the revised text.	
165	3.3 Design of ensemble projections	What are the possible implications of using the same fresh water flux adjustments in each of the ensemble members?	Dr Ed Hawkins	We don't think it's possible to speculate meaningfully on this, unfortunately, without running additional simulations. If we were to apply member-specific *patterns* of fresh water flux adjustment (not possible because we didn't have time to run member-specific calibration simulations), we would expect to narrow the range of regional salinity biases, and hence perhaps reduce the range of AMOC outcomes. However, it's less clear whether applying member-specific changes to the *global average* FW flux adjustment, without varying the regional patterns, would have made much difference to AMOC. We are looking at cross-ensemble variations in simulated P-E, surface salinity and AMOC, which will be reported in Yamazaki et al (in prep.). We have added a sentence pointing this out, in the second para of the spin-up section. If this analysis enables us to say anythihg more about the impact of our water flux adjustment strategy, we will include this in the paper.	
166	3.3 Design of ensemble projections	The 0.2C drift is not small, and could potentially be larger in the perturbed ensemble members. I would suggest some attempts to correct for this issue, or provide additional guidance.	Dr Ed Hawkins	We agree that it could potentially be important to correct for drifts, but we cannot do this until we have control runs for each of the GC3.05-PPE members. We plan to run these (see comment #83). The original control run quoted in the report is not ideal, because it uses present-day land use, as prescribed in the initial spin-up. Our new control runs will persist 1897 land use values to better mirror the historical runs, and so the historical runs will have an initial adjustment to this mismatch in land use. But this original control run is now about 190 years long and still somewhat informative. The text on the drift has been updated. Availability of more data and further analysis now indicates that the control run has a global warming of about 0.2K in the first 30-50 years (due to warming largely in Southern Ocean but also the tropics), and then settles down after that. This happens in both the standard historic run and the standard control run, and explains a good fraction of the warming seen in the first 40 years of the historical standard run. In the Southern Ocean, it is not obviously a drift though. Inspection of the 190 years of Antarctic sea ice outputs suggests this could be variability of about 30% in sea-ice fraction that occurs on a long (roughly centennial) time scale, due to convection in open water. This is a known model problem that happens in our CMIP6 models too (Ridley, pers comm). The Arctic sea ice in the control is largely stable, though does show decadal variability. That the adjustment seems largely confined to the early part of the historic runs means that drifts should not influence the anomalies relative to 1981-2000.	
168	3.3 Design of ensemble projections	Given the emphasis on plausibility, the retention of this cooling simulation implies that this is considered plausible?	Dr Ed Hawkins	We assessed it as plausible, based on the "two strikes and you're out" criterion that we used for this part of the assessment. In this sense, we adopted a practical view that none of the simulations could be expected to simulate well every single variable that users might want to look at. However, users do not have to use all the ensemble members if they think one or more members is/are biased to an extent that they deem unacceptable for their application. This should be stressed in the user guidance. These statements apply to PPE and CMIP5 members equally.	

169	3.3 Design of ensemble projections	How consistent is the selection of the CMIP5 models and the PPE models? Why were the same criteria not used? If exactly the same criteria had been used for both, would the selection have been the same?	Dr Ed Hawkins	Similar principles (of plausibility and diversity, and use of a mix of quantitative and qualitative criteria) were used. However, we did not use exactly the same set of diagnostics, mainly due to resource and timing constraints. We decided that the most efficient way to perform the CMIP5 selection was to rely as much as possible on published analysis, which gave an advantage of being able to access a range of evidence, but meant that we were reliant on methodologies (such as the Sanderson et al. near-neighbour analysis or the Perez et al weather typing) that we would not be able to replicate in-house. However, sections 3.4 and 3.5 present a range of common analyses carried out on both ensembles, so readers can assess their relative performance in a clean fashion.	
170	3.4 Evaluation of combined Strand 2 ensemble	Given the importance of the NAO for UK climate, is this significant bias not enough to rule out these simulations?	Dr Ed Hawkins	We don't think so. Our general principle, as explained in #168, is not to rule out simulations for a poor simulation of one particular variable. Another consideration was whether specific ensemble members were outliers from the rest of the ensemble. Although, for example, GFDL-ESM2G is significantly higher than the observed value, all members are biased high in their NAO amplitude, so this makes it harder to argue that GFDL-ESM2G should qualify for an outright red flag. Users always have the option to account for this information on NAO amplitude, or other diagnostics in the science report, on an application-specific basis.	
172	3.4 Evaluation of combined Strand 2 ensemble	Fig 3.11 is concerning. There is some discussion of implications for long-term warming, but no discussion of near- and mid-term warming, which is likely more related to the aerosol focing and high climate sensitivity issues. This needs to be discussed here as the near- and mid-term is arguably more important for adaptation decisions. The long-term drift identified earlier is also important here - if that is subtracted then the comparison is worse.	Dr Ed Hawkins	This comment was inadvertently overlooked in the initial revision. See right for response following Moderation Meeting.	At the Moderation Meeting, use of a 1981-2000 baseline was suggested for this Figure. Following subsequent further discussion, during which this was considered, it was decided to keep the original 1901-50 baseline, but add the available control simulation of the standard GC3.05-PPE member as an illustration of the potential effects of climate drift. Details of how the model results were masked and blended (to match better the nature of the observed data) were also added. Text was revised, including the following comment relating to near-term warming: " However, the mid-20th century cooling may be more informative regarding near-term future GMSTchanges, than it is about longer-term levels of warming. Assuming that thehistorical aerosol forcing in GC3.05-PPE is too strong in most members, it alsofollows that the effects of future reductions in aerosol forcing, which may be expected to enhance projected warming trends during the next few decades (Raes and Seinfeld, 2009), are likely to be overestimated."
174	3.5 Projected global changes in combined Strand 2 ensemble	Are the projections for crop yield thresholds bias corrected?	Dr Ed Hawkins	No. The figure only shows projections of simulated temperature and precip anomales for growing easons, not predictions of crop yield per se. We quote growth thresholds only to illustrate how the model results might tbe used in this application. We added: "Note that in a practical application of this type, it would be important to consider carefully how to bias correct the model output for use in crop projections, given especially the importance of thresholds in the calculations (Hawkins et al., 2013)."	
175	4.4 Evaluation of regional simulations for 1981-00	Figure 4.3 lacks ERAI-RCM-STD. The reader want to know to what extent the used RCM is an appropriate RCM or not which requires knowledge of how it behaves when forced by perfect-boundary conditions. It also requires that the model is appropriate for simulating climate change - this could to some degree be assessed by looking at a different (warmer) climate in todays situation (for instance can the RCM simulate a drier and warmer summer climate over continental Europe in todays situation we would be more confident in trusting it on saying stg on warm dry summers over the UK in the future) - therefore this section would also gain from figures showing the situation for all of Europe (cf. the pair of Fig 4.1 and Fig 3.12).	Prof Dr Erik Kjellström	Thanks for the suggestion. We agree that use of ERAI-RCM-STD, especially when combined with analysis of RCM-STD, the corresponding GCM-driven simulation, is a good idea. This allows us to precede evaluation of the GCM-driven simulations (which are the main priority for users) with an assessment of the regionalisation skill of the RCM, and to perform a clean comparison between the effects of driving model and downscaling biases, as done in Fig. 4.1. We have added a sentence to the end of section 4.3 to emphasise this. We also agree with the idea of extending such comparisons to extremes metrics (as we have done in Fig. 4.4), and will consider including more examples of these in the relevant specialist journal paper that is being written on the Strand 3 runs. However, Figure 4.3 focuses on biases in the ensemble-mean of the RCM simulations, so we do not think it would be appropriate to add ERAI-RCM-STD results to these particular plots. This is because the differences in bias would be more complicated to interpret than those of Fig. 4.1. Since it is the GCM-driven regional model simulations that will be used in impacts studies, we have prioritised assessment of these in the present report. The idea of extending Fig. 4.3 to cover Europe is also something that we feel would be best covered in a specialist journal paper. While considering present-day conditions in parts of Europe as an analogue for potential future UK changes is an interesting idea, it comes of course with scientific caveats. These would relate to differing physical influences that might drive spatial contrasts in present climate, as opposed to location-specific future changes. So we feel that a discussion of this nature would need to be supported by a more extensive analysis and discussion than is feasible in this report.	
176	4.4 Evaluation of regional simulations for 1981-00	How were the percentiles determined? Are they empirically taken from the data or are they taken from any fit of an analytical function to the data?	Prof Dr Erik Kjellström	They were calculated empirically from the data. This is now stated in the caption to Fig. 4.3a.	
177	4.4 Evaluation of regional simulations for 1981-00	How are urban effects represented in the RCM? Or, should it rather read "the lack of representation of urban effects in the model"?	Prof Dr Erik Kjellström	The RCM does represent urban effects, as one of the "tiles" in its land surface scheme. However, the forthcoming 2.2km simulations will include a more sophisticated scheme, involving the use of two tiles to represent roofs and street canyons, in which surface parameters are determined from the morphology and materials proerties of the relevant city (Porson et al., 2010). Text reworded accordingly.	

179	4.5 Projected changes in the Strand 3 ensemble	I lack information about statistical significance for the changes in this chapter. Did you assess it? If not, why???	Prof Dr Erik Kjellström	As the Strand 3 product is an ensemble, we thought it preferable (here as in section 5) to show the changes as 10-50-90 percentile maps. We think this is more flexible than some specific significance test, as it allows readers to see the range of results, as well as providing information that can support statements of statistical significance. For example, where temperature responses show warming at the 10th percentile, coupled with larger levels of warming at the 50th and 90th percentiles, this is equivalent to a statement that: "according to this ensemble, there is at least a 90% chance that location x will experience warming by 2061-80 relative to 1981-00 under RCP8.5 emissions." Similar statements can be made for any variable and location for which the 10, 50 and 90 percentile changes are the same sign, as is the case in many regions for winter precipitation (increases), summer precipitation (decreases), and the extremes metrics that we show. In cases where the responses at the 10th and 90th percentiles are of opposite sign, this shows that there is less confidence in the sign of the changes in the ensemble. We have, however, avoided using this sort of probabilistic language in the report text. This is because the results are specific to the relevant ensemble (here RCM-PPE), which is not designed to be interpreted probabilistically (see #50, in particular). Hence we prefer to avoid this kind of presentation, to avoid the risk of inadvertently encouraging users to interpret these conditional results as statements of probability in some broader sense.	
181	4.5 Projected changes in the Strand 3 ensemble	These increases of "up to 10%" are hardly discernible in the figure. Are such small changes statistically significant in some way?	Prof Dr Erik Kjellström	We agree that these changes are unlikely to be statistically significant. We changed the sentence to read: "The only exceptions are some coastal regions of Northern Scotland, where the 90th percentile shows the potential for some small increases."	
184	Tables	E-OBS is around in a large number of versions, version number is lacking here.	Prof Dr Erik Kjellström	We used version 14 for precipitation and version 15 for surface air temperature. Table D.3 has been amended to record this.	
185	Figures	Figure 4.4 ERA-Interim is hardly legible - I have troubles distinguishing the colors from each other in this figure. Also, the nature of the curves with them all being relatively similar makes it difficult to follow statements in the text saying that one of the curves are in better agreement than another. Could you consider showing also relative differences between the PDFs here in some way?	Prof Dr Erik Kjellström	Done. We kept the 12km RCM row that shows absolute values of the obs and ensemble values, so readers can see the general shape of the contributions. Then showed the same results as differences from obs, and corresponding GC3.05-PPE differences undereneath. We got rid of the RCM @60km row, as these look similar to the 12km results. Plotting as differences hopefully makes it much easier to see the variations in behaviour between the different simulations - thanks for the suggestion. Text revised accordingly.	
186	3 Strand 2: A new ensemble of global climate realisations from 1990-2100	I don't have many specific comments because, overall, the work is of an extremely high quality. My main concern is more from a user perspective. The PPE and CMIP5-13 ensembles essentially quantify uncertainty from two different perspectives - one parametric and emissions driven, the other structural. Neither can be viewed as having objectively sampled these sources of uncertainty. That is, in both cases, from a statistical perspective, neither the population of plausible representations of our climate nor the sampling scheme that draws from that population, is known. This makes it difficult to advise users on how to interpret, and accord weight to, differences between the two ensembles – and it makes it very difficult also to provide well founded advice on how to merge information from the two ensembles. In the case of the PPE, the starting point is the STD model, which has previously been selected somehow from the population of plausible models. This choice itself is not random – clearly a great deal of expert judgement is involved in the development and tuning process to produce a "best" representation of the climate system. Conditional on this choice, the PPE could be thought of as a sample of perturbations where the sampling scheme that produces the perturbations is more or less known, except that expert judgement again comes into play in the many decisions that play a role in determining which of the 25 initial ensemble members are retained. In the course of doing this, some implicit insight is obtained concerning the effect of tuning during the model development process via the positioning of the STD model relative to the other PPE members in the evaluation of the various performance metrics. This suggests that perhaps selection bias may have been a factor in the determination of the STD model. In the case of the CMIP5-13 ensemble, the starting point is the much larger CMIP5 ensemble of opportunity. The CMIP5 ensemble samples structural differences in the representation of the climate system, but we do now know the sampling strategy or the population from which that strategy draws realizations. Also, each model would have been subject to a development and tuning process that one might suspect would lead to some kind of selection bias. A complication is that the nature of this bias is likely different for each	Dr Francis Zwiers	We agree that there is no known distribution for the PPE and CMIP5-13 runs in Strand 2, and hence these ensembles do not quantify uncertainty and will not provide an estimate of risk - that is the role of the probabilities in LS1. Instead the LS2/LS3 products provide plausible realisations of what a future UK climate might look like, and its purposes are outlined in several places in the report, for example in section 3.1. There were several examples of how flexible data sets provided by raw model output in UKCP09 were used to initially investigate climate impacts, before the question of risk was raised. The Strand 2 and 3 results in UKCP18 can also help with a storyline approach, to understand better (via physical explanations) why the climate is changing. The Australian, Swiss, and Dutch projections are all based on CMIP5 runs, so many of the tools they use could in principle apply to Strand 2 and 3 runs.	
187	3 Strand 2: A new ensemble of global climate realisations from 1990-2100	The other general comment that I have is that the use of a 20-year reference period (in this case 1981-2000) seems uncomfortably short to me – even if this has been IPCC practice for the last couple of assessment reports. The general problem can be seen in Figure 3.19 – all curves are required to pass through zero between 1981 and 2000 with the result that there is apparently less uncertainty in representing climate change during this period, than at other times, with the fan of uncertainty widening in both directions as distance in time from the reference period increases. A longer reference period (30-years, the entire historical period, or the entire 20th and 21st centuries) would impose weaker constraints on variation within the reference period and thus avoid the sense that models reproduce the recent 20-years used for the reference period best.	Dr Francis Zwiers	The reference period was agreed with users, so we have to use it. We agree that use of a 20-year period has these limitations, but a 30 year period would also be open to the criticism of a greater degree of non-stationarity, due to climate change trends. We do use 1901-50 in fig 3.11, which allows bases in simulating historical changse to be seen more clearly.	
190	3.3 Design of ensemble projections	Note that this kind of sampling can lead to aliasing problems	Dr Francis Zwiers	We don't understand this comment.	

194	3.3 Design of ensemble projections	why 3.7 and not some other threshold; what determined the choice of threshold?	Dr Francis Zwiers	The method contains some parameters like this threshold, which can be adjusted. This is the main adjustment parameter. We only had so much supercomputing resource to run potential candidates sampling a 47-dimensional model parameter space, and we ruled variants out using a large number of metrics. When you do that, you have no way of knowing how many of the original runs will remain after the filtering. From a previous study (Karmalkar et al., submitted), we know that somewhere in the range 2-4.5 works well, from using CMIP5 as a benchmark. Adjusting the threshold parameter controls how many variants survive the filtering. If the threshold is too low, then we rule out too many (see next comment) and do not have enough variants to run in coupled mode, noting that we have to allow for some coupled runs to drop out due to unpredictable coupled behaviour (AMOC collapse, instabilities etc). But we do not want the threshold too high as this lets too many runs through, raising the risk of generating implausible coupled members. So 3.7 was chosen as it let 49 variants through. We felt this was the right balance, letting enough through to allow us to pick a subset of the 25 most diverse members.	
195	3.3 Design of ensemble projections	how sensitive is this number to the choice of threshold?	Dr Francis Zwiers	A choice of threshold=3.0 rather than 3.7 let 17 rather than 49 through. Threshold=4.0 let 52 through. So as you reduce the threshold, the number retained seemed to drop off more quickly than if you increase it.	
200	3.3 Design of ensemble projections	these filtering steps presumably should be understood as further conditioning the PPE on the observed state – so the quality of the observations, and perhaps the particular realization of low frequency variability that has been experienced over the past century, become important.	Dr Francis Zwiers	To a large extent, the PPE could be viewed that way, although some of the variants that were ruled out were due to numerical instabilities (a large fraction in 5-day forecasts, a smaller fraction in the 5-year runs, and two in coupled runs). The filtering for the 5-day and 5-year runs was based largely on a quantitative approach, and observational uncertainty (from both "measurement" error and internal variability) were accounted for. The filtering based on historical performance was qualitative. We agree that the two sources of observational uncertainty are important and should be considered. If we had used a quantitative approach in the final filtering steps, we would have had to account for them explicitly. But in the qualitative filtering, their inclusion is implicit, and we only ruled out members that showed unacceptable performance to a level where observational uncertainty, internal variability, and some tolerance on structural errors still would not have saved that member. For example, the NH cooling assessed in the final step is not consistent with the IPCC AR5 assessment, which would have considered the sources of observational uncertainty that you mention.	
201	3.4 Evaluation of combined Strand 2 ensemble	the report deals with the cold bias in a factual way, but it doesn't seem to deal with the implications a great deal. I think it does imply that the PPE, and the spread amongst PPE members, represents quite a different sample from model space than the CMIP5-13 sample. How should users make use of this information, and how should it affect their choices of what global model results to use for their impacts and adaptation assessments?	Dr Francis Zwiers	Yes, this is an important point that needs to be made clear in the user guidance. There are two aspects to this. First, we should recommend that Strand 2 output should be evaluated on a case-by-case basis, and that users should not use members with large biases if they feel that these are having a significant effect on the projected climate change. Second, users need to realise that in some cases, particularly for temperature variables, the PPE and CMIP5 will represent different samples, and that Strand 2 should not be considered as a distribution that can be randomly sampled to provide a few representative members. Users will need to inspect the Strand 2 members and select the members that are most useful for their application. For example, if a user is interested in seeing what a high warming scenario looks like, they deliberately pick one of the PPE members.	
207	3.4 Evaluation of combined Strand 2 ensemble	given that precipitation is so bad, is it usable at all, even after bias correction / downscaling?	Dr Francis Zwiers	We now cite the need to consider the origins of large errors on a process-basis, and refer back to the discussion added in response to #208.	
208	3.4 Evaluation of combined Strand 2 ensemble	Figure 3.13: I can see the attraction of showing variability biases in physical units, but I nevertheless find it awkward to display the difference in variability in this way. The "natural" comparison, to which an F-test can be readily applied, is to plot variance ratios (e.g., dividing the pooled variance across the CMIP5-13 ensemble by the observed variance). Ratios different from 1 tell the tale about biases, and a zeroth order comparison to guide how seriously to take a given value can be made by comparing the calculate ratios with critical values from the F distribution. This immediately also points out a weakness of assessments against 20-year periods. Assuming no excess low-frequency variability (ie., that annual seasonal means are independent and identically distributed), the observed variance has only 19 degrees of freedom. This means that the critical values for determining whether a simulated variance is significantly different from the observed variance are relatively large. The 5- and 95% critical values assuming 13x19 degrees of freedom in the numerator (pooling variance estimates from 13 CMIP5 models) and 19 degrees of freedom in the denominator are 0.614 and 1.904 respectively. Relaxing the assumption of iid interannual variability to account for serial dependence would move the critical values even further from 1. Where do the values shown for the CMIP5 models fall within this range? Note also that the power to detect departures from the null hypothesis of the equality of variances is low – large samples are required to reliably detect, for example, a doubling of variance. A longer base period, allowing a few more observational degrees of freedom, would help – but overall, assessing differences in variability is difficult.	Dr Francis Zwiers	We agree that a ratio would be more naturally suited to a significant test, but our propose here is only to give a simple comparative indication of the main biases in the simulations. Given the uncertainties in the observations (particularly for precipitation, see #67), plus the use of a short period (chosen for consistency with other diagnostics, and to allow the RCM-PPE runs to be compared with Strand 2 on a like-for-like basis), we don't think a formal significance test would be all that useful.	
210	Whole report	In terms of the observational constraints: its very interesting to use CO2 - is there any chance that agreements could be spurious, ie that there are missing factors in the carbon cycle model that would lead to model versions agreeing with data that are not inherently better? I realize there is a paper but as the constraint is strong I recommend discussing this here briefly.	Prof Gabi Hegerl	Indeed there are several processes that drive the net carbon cycle feedback, hence there is no guarantee that any given model that gets the past right will necessarily get the future right. This is, of course, why we sample multiple possible futures in the pdfs, to get a set of alternative outcomes conitioned on the obs. Also, this multiple drivers point doesn't just apply to carbon cycle feedback. It applies to historical GMST changes (models could get thiese right due to a wrong balance of physical feedbacks, aerosol forcing and ocean heat uptake), or indeed any observable. We have covered this point in our revision of the text in Appendix C (now in main text), but in a general way, rather than focusing specifically on carbon cycle feedback.	
212	Whole report	The observational constraints are limited to temperature. Is there any danger that use of a different variable (pcp; slp?) would have given different results? It is probably not feasible to use a different variable but a brief discussion of this would be helpful.	Prof Gabi Hegerl	The constraints are not limited to temperature. The recent climatology constraints (CLIM) include slp, precip and a set of other variables commonly used to evaluate climate models (Table B.1), and the historical changes constraints include CO2 as well as surface temperature and OHC metrics. We acknowldege that the description in the main text was rather terse, so we have moved Appendix B to the main text and expanded the introduction to the observational constraints (section 2.2, Stage 2c) so that the set of observablves is more visible.	

214	2.1 Overview	for what variables is HadCM3 competitive? Probably not all!	Prof Gabi Hegerl	Naturally, the relative performance of HadCM3 with respect to other models is different for different variables, as is the case for any model. Even the new GC3.05 model does not perform well at everything, as is made very clear in section 3.4 ! We don't think it would be necessary or useful to give a detailed list of how HadCM3 breaks down against CMIP5 models on a variable-by variable basis, as the judgement of credibility was, and should be, a broadly-based one. However, we have added a footnote, listing the assessment criteria that were used. This reads: "The assessment included quantitative evaluation of spatial fields of multi-year climatological averages of surface temperature, precipitation, pressure at mean sea level, net radiation, net cloud radiative effect and its shortwave and longwave components, planetary albedo, and atmospheric temperature and zonal wind at 200hPa and 850hPa. Also included was qualitative evaluation of the NAO, historical trends in Arctic sea-ice extent, the North Atlantic storm track and the frequency of atmospheric blocking events."	
215	2.1 Overview	how does this step compare to the one using observational constraints? Are the observations used twice?	Prof Gabi Hegerl	The screening of ESPPE members was a simple qualitative "history matching" exercise using a couple of global average quantities and knowledge of observed vegetation distributions. The exercise was designed only to remove members with large biases, that would have received negligible weight in any formal likelihood calculation. In any case, the constraints in the probabilistic calculations were calculated from *emulated* estimates of the constraining observables across the entire parameter space, not from the ESPPE runs. So, to all intents and puroposes, there was no double-counting. Text revised to explain the screening a bit more fully: " This was reduced to 57 members, following a simple screening based on historical performance to remove members with substantial biases in planetary net radiation, global surface temperature or simulated vegetation distributions (Lambert et al., 2013). Murphy et al. (2014) compared the performance of the surviving 57 members against CMIP5 earth system models for a few key aspects of historical climate, finding similar levels of skill in both ensembles ." Footnote: The comparison with CMIP5 models considered recent climatological patterns of surface air temperature, precipitation and pressure at mean sea-level, plus simulations of changes in GMST during the 20th century.	
220	Figures	what does this figure look like for the posteriors?	Prof Gabi Hegerl	The impact of the observational constraints is to narrow the posterior distributions. The degree of narrowing is typically modest for UK variables, and somewhat larger (at the top ends of the range) for global temperature metrics. Examples of this are shown in Appendix B, Fig. B1, in support of the subsequent material in Stage 2c. However, we don't show the impact of observations in Figure 2.3, because its specific purpose is to illustrate aspects of the methodology for constructing prior distributions. Another reason is that the variables of Fig. 2.3 are idealised metrics (ECS and normalised UK changes), whereas in Appendix B we think it more relevant to UKCP18 to show examples of its actual outputs. While posterior distributions of ECS (in particular) are (of course !) interesting to the wider climate science community, to show these here would require considerable supporting discussion and context, and would be a diversion from the purpose of section 2.2. We would therefore prefer to defer such discussions to subsequent specialist journal papers.	
222	Figures	figure 2.8 scotland precipitation bit: I agree that a wider tail captures uncertainty better, but I am surprised its wide on both ends. Do we really think there is a serious probability that scotland precipitation will decrease by 40%? Where does that come from? the observations if anything suggest a strong change.. (eg Min et al. 2008 although uncertainties are substantial). A discussion in the text about aspects that changed such as this one would be useful. if its considered less likely in practice given other things we know it might be useful to caution this.	Prof Gabi Hegerl	Note that the possibility of 40% negative anomalies is shown only in the distribution of *seasonal* anomalies in fig. 2.8, and arises from the effects of internal variability. In the distribution of *20-year mean* changes, the lower limit of possible changes is put at ~-20%, though the bulk of the pdf suggests positive changes, as we would expect. Fig. 5.1 shows that the characteristic climate change response of Scotland precip in winter has a cumulative probability of around 10% of being negative. For any given 20 year-mean period, low frequency variability (e.g. as shown for England in Fig. 5.4) will likely boost that probability somewhat, so we think the blue curve in Fig. 2.8 is defensible, albeit that the balance of probability is certainly for an increase.	
224		It would be useful to understand the differences between GC3.05 and GC3.1	Prof Jim Hall	Appendix D had been expanded to give a full list of differences between GC3.05 and GC3.1. A forward reference to this information has been added to section 1.2.	
225		It needs to be clear that these pdfs are conditional i.e. on emissions and several methodological choices. This sentence seems to imply that they are unconditional pdfs	Prof Jim Hall	Added "conditional" to the relevant sentence. Also, this point is now made earlier in the report, by adding an extensive discussion in section 2.1 See #95.	
226		It's disappointing that the number for RCM runs for downscaling could not have been increased for UKCP18. 11 members is a small sample given the dimensionality of the parameter space	Prof Jim Hall	This comment was made in the context of Strand 1. Strand 3 does provide a new RCM ensemble, though it is still only 12 members. The size of the ensemble was restricted by the computational resources and the project time scale. However, we would also note that the *design* of the ensemble is important too. given that the Strand 3 members all simulate a strong future summer drying, for example, it's not clear that adding (say) and extra 20 PPE members would have been that useful. As noted in #72, it may be more fruitful to point users towards alternative ensembles such as EuroCordex, in order to obtain a more divers set of plausible downscaling scenarios.	
231	2 Strand 1: Probabilistic projections	The reason for the lower median temperature projections looks like it is understood (CO2 constraint). However, it is really not clear why the precip uncertainty goes up in Figure 2.6. I suppose I was expecting that uncertainties should reduce with the improvement in the implementation of the method and inclusion of additional observational constraints. Is this something that is specific to the UK or e.g. also a feature of global mean precip (if that is a variable that is output)? Presumably it translates also to PDFs at the grid scale? If so, some commentary is required (as indicated by the placeholder text).	Prof Mat Collins	We agree, having flagged the wet tail in Fig. 2.6 as an issue to understand. Further investigation has revealed that it was an unrealistic feature resulting from use of a logarithmic transform applied to the preciptiation variable during the statistical calculations (prior to back-transforming into % changes in the final output). This transform was a legacy of UKCP09 and Sexton and Harris (2015), and was a useful way of avoiding statistical generation of negative precipitation amounts. However, it was found that in cases where: (a) a very strong long-term drying is predicted, (b) internal variability is high (e.g. for monthly variables in small regions, not considered previously), and (c) the variability is correlated with the long-term response, the log transform could give unealistically high variability at the wet end of the distribution. We therefore ditched the log transform, and repeated the statistical calculation without transformation, in units of % anomalies, using clipping to remove any instances of negative anomalies exceeding -100 %. This fixed the wet tail issue, while giving cumulative probabilities for dry-end thresholds similar to those obtained with the log transform. Figure 2.6 will be updated in the final version.	
235	3 Strand 2: A new ensemble of global climate realisations from 1990-2100	From a scientific point of view, the lack of analysis and understanding of why the ensemble exhibits those interesting/unusual features is required. For example, it would seem essential to provide some more in-depth analysis of why the ensemble exhibits such high levels of global warming in comparison with CMIP5 and the strand 1 scenarios. There is one figure on changes in cloud radiative effect but I hear rumours about stratospheric ozone ending up below the tropopause (a classic HadCM3 propblem too) and also feedbacks associated with using the easy aerosol scheme. Perhaps also snow albedo feedbacks (see below)? I am sure the team ran out of time to do the sort of in-depth analysis required, but perhaps there is some additional work that has been done subsequent to the writing of this report that can be added.	Prof Mat Collins	We agree that more understanding is needed, but provision of Strand 2 was always contingent on the idea that it would be a new and somewhat experimental product, the properties, strengths and weaknesses would be learned about over time, through community analysis. The current position regarding the high warming of GC3.05-PPE is summarised in #49. We have a work-in-progress plot looking at albedo feedback, and in fact its global strength in the PPE is not larger than in CMIP5. There is some evidence that the cold bias has caused a regional increase in albedo feebdack over parts of Eurasia. We will put this in the final version of the report if time permits.	The plot of albedo feedback was included.

236	3 Strand 2: A new ensemble of global climate realisations from 1990-2100	The central issue, as identified by the UKCP18 team, is one of assessing the credibility of this ensemble w.r.t. CMIP5 and strand 1. One way of looking at the ensemble is that it is based on a better model, with higher resolution, improved processes and reduced mean climate biases. Hence the high sensitivity ensemble is more credible. The other side of the coin is that the ensemble lies (well) outside the likely range of warming as assessed in e.g. IPCC AR5 and thus the ensemble us much less credible. Maybe the team are not willing to come down on one or other of these two arguments but there does need to be some discussion of the issues and, ideally, more evidence. Ask yourself what would the AR6 authors make of this strand 2 ensemble? Might it influence their thinking on the AR6 assessed range of future warming, or would they dismiss it?	Prof Mat Collins	Although we don't have ECS values for the new PPE members yet, the emerging evidence discussed in #49 suggests values around 5 degrees. Irrespective of the values, we don't think it necessarily follows that either the results are more credible simply because the model is new, or of a high resolution, or that the emergence of values beyond the AR5 likely range (though perhaps not the very likely range - see https://www.ipcc.ch/publications_and_data/ar4/wg1/en/ch10s10-5.html#box-10-2) should necessarily imply a lack of credibility. Only detailed analysis in slower time can determine credibility, and even then, the answer may well be that (as with most climate models) it is difficult to prove categorically whether the simulation of feedbacks is any more or less plausible than in other models. We don't see any obvious reason, for example, for AR6 authors to disimss the runs - the general level of performance seems competitive with CMIP5 models, while acknowledging that the simulations of historical GMST will attract attention, and require careful communication from a "beauty contest" point of view. The GC3.05-PPE runs explore the top 10% of the Strand 1 distribution of GMST outcomes, so we think it is reasonable to characterise them as providing high but plausible values of global warming, that can be useful for looking at high-end impacts.	
237	3 Strand 2: A new ensemble of global climate realisations from 1990-2100	In addition to the sensitivity issue, there seems to be some quite large biases in the strand 2 ensemble as evident in figures 3.8, 3.12 and 3.14. Although the picture is a bit confusing to me with the switching between mm/day and % for the precip biases. Does 3.8c indicate a dry bias in western Europe in DJF and 3.12b indicate a wet bias? (Suggest sticking to mm/day for this scientific document). Might also the cold winter continental temperatures contribute to a larger snow-albedo feedback? The authors suggest 'consideration of bias correction' but when the biases, particularly in winter, are so large, is this even sensible?	Prof Mat Collins	We used mm/day in Fig. 3.8 becasue it was a global map including arid regions, while % biases were used in Fig. 3.12 because many users like to work with % future changes, for bias correction purposes. As noted in #227, this report is aimed at expert users as well as the IPCC WG1 community. However, we agree that there do appear to be some differences in the sign of bias over Europe. It could perhaps relate to use of different verification datasets. We'll check, and revise the final version as needed. As noted in #235, there is some evidence that the cold bias may have enhanced regional albedo feedbacks. A discussion of the limitations of bias correction has been added. See #207.	The different signs of bias (over parts of northern and eastern Europe) are indeed due to differences between GPCP and E-OBS: GPCP is wetter, apart from in high-elevation regions. We have added text to the existing discussion of measurement caveats, to point this out.
240	5 Projections of future variability and change for the UK	Perhaps an adidtional figure ought to be included that shows absolute changes in a common time period? This normalised version obscures the differences between the different strands.	Prof Mat Collins	We agree. Figure 5.2 does this, for 2061-80 wrt 1981-00. We thought it better to show these full (non-normalised) changes as maps rather than as national averages, to convey more information and give a better sense of the spatial nature of the products. However, for England we do also show time series of the full temperature and precip changes in Figs. 5.3 and 5.4. So we think this point is already covered.	Following the Moderation Meeting, a version of Fig. 5.1 showing full changes was added.
241	5 Projections of future variability and change for the UK	These plots are quite useful for seeing the leading-order differences between the strands for UK climate. I wonder if additional plots for other key variables could be included in an appendix?	Prof Mat Collins	We see the Science Report as providing selected examples illustrating the types of information available from UKCP18, and how the Strands compare, rather than as the main resource for large collections of plots (see also #49). This is to keep the size of the report to manageable proportions, and avoid diluting the basic underpinning messages that we think readers need to take away. So, we would prefer not to add more of these plots to the report, but more examples will be added to the User Interface, after launch. We've added a link to the text.	
242	3.3 Design of ensemble projections	lines 10-11: The sampling of diversity is also subject to the constraint of the number of ensemble members used	Prof Ted Shepherd	Revised sentence to read: "Here,the aim is to ensure that the broadest possible range of future outcomes is sampled, subject to the scientific constraint of ensuring plausibility in the selected members, and the technological constraint on ensemble size imposed by computational expense (see above)."	
244	3.3 Design of ensemble projections	line 24: Are 5-year averages really sufficient to determine such biases? The typical atmospheric dynamicist would be inclined to think not, so some discussion is surely warranted here.	Prof Ted Shepherd	Added a footnote to address this: "Ideally, longer simulations would have been preferred, in order to achieve better sampling of the effects of internal atmospheric variability in the AMIP runs. This was prevented by limitations in HPC resources. However, we found (as reported in section 3.4b) that parallel AMIP simulations at N216 and N96 atmosphere resolution gave consistent results for model biases, suggesting that the five-year simulations were long enough to give a reasonable estimate of climatological errors in the 25 members selected for use in coupled simulations."	
245	3.4 Evaluation of combined Strand 2 ensemble	line 10: Perhaps the very large relative biases in precip are because the absolute amounts are not very large in these regions, so are actually not so meaningful?	Prof Ted Shepherd	Added: "Note, however, that since summer precipitation levels are typically low in these regions, the large percentage biases do not represent especially large errors in absolute precipitation amounts."	
246	3.5 Projected global changes in combined Strand 2 ensemble	lines 6-8: Isn't the difference between Strand 2 and obs In the current century mainly because of the cold bias in Strand 2 during the reference period?	Prof Ted Shepherd	Possibly, but additional single-forcing experiments would be needed to confirm this. The hiatus in observed warming is already mentioned as one potential factor. We added the following sentence: "In GC3.05-PPE, another contribution may arise from warming due to the reduction in global sulphur dioxide emissions after 1980, given the strong aerosol forcing simulated in the ensemble (sections 3.3d and 3.4a)."	
247	3.5 Projected global changes in combined Strand 2 ensemble	line 5: It could be emphasized that Strand 2 should not be considered as (emissions-conditional) predictions, in the same way that Strand 1 can. As it turns out, GC3.05-PPE may be especially useful for stress-testing temperature-sensitive impacts. I realize you say this later, but it is important to ensure this figure is not taken out of context!	Prof Ted Shepherd	We think the text above (the lines referenced here) makes the point that Strand 2 results are designed to provide possible realisations, rather a set of results that can be interpreted in terms of risk or probability of different outcomes. We have added : ".For example, the GC3.05-PPE runs may be particularly useful for stress-testing impacts sensitive to high levels of future warming (see also discussion in section 5.2)."	
248	5.1 Comparison of projections from Strands 1-3	line 2: I am concerned that it is misleading to use a baseline of 20 years for such a figure. Since the filtered data only shows variations slower than 20 years, you are defining a baseline by essentially one data point. The apparent discrepancies come from the unusual observational values during the baseline period. My suggestion would be to drop the figure since it is difficult to interpret. Anyway, I believe that is not statistical best practice to show smoothed time series. (Block averages are fine, of course.)	Prof Ted Shepherd	We agree, of course, that any given 20 year period is influenced by low frequency variability as well as the underlying climatology of the period. However, the reality is that the UKCP18 stndard baseline has been set (in consulation with users) as 1981-00, and it is important to recognise that there are competing drivers for that choice, such as the desire to have a reasonably contemporary period, and to avoid a longer period that might show larger underlying climate change trends, creating problems of non-stationarity. Given that the choice has been made, we think it is better to include a figure such as 5.4 (also 2.9), and then expose the issues to aid understanding, rather than to remove the figure (and hence fail to advise users). We added this sentence, to emphasise the point: "This discussion also illustrates that observed 20-year averages for a given period (such as the UKCP18 baseline of 1981-2000), can be significantly influenced by the dominant contemporaneous phases of low frequency climate variability, as well as by the true underlying climatological state." Regarding filtering, our understanding is that simple averages are not an efficient way of removing the effects of high frequency variability (our specific aim in Fig. 5.4), and the cut-off properties are sharper when a specific filter is used. In the Butterworth filter that we use, the values obtained for a given period like 1981-2000 are influenced by values outside as well as within the period, in contast to the use of straight averages. So, we would argue that filtered values are somewhat less affected by the "one point" issue.	
249	5.1 Comparison of projections from Strands 1-3	lines 12-17: It could be mentioned that this pattern is consistent with CMIP5 projections as a whole, as shown by Zappa et al. (2013)	Prof Ted Shepherd	Done.	
252	3.4 Evaluation of combined Strand 2 ensemble	I don’t see much evidence, based on Fig 3.14, of differences between E-OBS and NCIC that can be attributed to different numbers of stations. Do the numbers, 138 and 3000, refer to stations in the UK or Europe as a whole?		The station numbers we quote are indeed for the UK. Revised text to confirm this, and explain that the effects of lower station density in E-OBS are particularly apparent in high-elevation regions, in which E-OBS consistently shows lower values than NCIC. This is shown in Hofstra et al. (2009), the paper we cite as supporting evidence. Fig. 3.14 only shows UK averages, so isn't designed to provide proof of the effects of higher density in NCIC. However, the region-specific biases will of course contribute to the UK average differences between the two daasets, which is the point we're making.	

254	3.4 Evaluation of combined Strand 2 ensemble	FIgure 3.18, why does CMIP5 start at about 1950 in these plots? What can be said about the quality of the EMSLP dataset – how homogeneous is it, and how does it compare with MSLP from 20th Century Reanalysis (20CR)? The 20CR is useful for this kind of question because it is an ensemble product. Comparison between ensemble members (there are 56) tells you something about the strength of the observational constraint on analysed MSLP at any time in the history of the reanalysis, which in turn would give some sense of whether variation in WT1 frequency, for example, is affected by the change in the configuration of the surface pressure observing network over time.		For CMIP5, we exploited weather typing analysis that had already been performed by a colleague (from 1951 onwards), but did not have resources to extend the analysis back to 1900. We have revised the text to note that the CMIP5 data does not go back to 1900. Unfortunately, it is not feasible to analyse the (interesting !) 20CR dataset in this way, on the project delivery timescale. However, the validation results we quote relate to typical behaviour over the whole 20th century, so we don't think it matters too much that there is a degree of uncertainty attached to results from individual years. Regarding longer term variations in the observing network, Ansell et al (2006) conclude that EMULATE is robust post-1880, because the dataset is derived by blending station data with chart analyses based on thousands of observations. They also show that the time series of the NAO compares well with the NAO diagnosed from its defining stations, so we think the data is reliable.	
257	3.4 Evaluation of combined Strand 2 ensemble	how is “plausibility” assessed?		Each Strand is designed to produce some range of outcomes that cannot be ruled out on the basis of current knowledge. This is what we mean be "plausible". Each range is conditioned on the sets of models, and evaluation criteria, used in their development. This includes Strand 1, which typically gives the broadest ranges, because the sampling contained within it is more intensive than for Strands 2 and 3. Therefore, it can be used to show the limits of the sampling of outcomes in the other Strands, which is what the relevant piece of text is saying. These principles are explained in many parts of the report, so we think the meaning of plausible should be clear to readers at this point in the text, without requiring a lenghty restatement of the above context.	
261	3.4 Evaluation of combined Strand 2 ensemble	this point could have been usefully made MUCH sooner.		We've moved this sentence to where Fig 3.21 is first introduced.	

Bias correction

ID	Section	Comment	Reviewer initials	MO response	Post Moderation Meeting Met Office comments
3	Table	These are all addiive adjustments. Might want to discuss relative adjustments used for precipitation.	Prof Ted Shepherd	Relative adjustments have now been added to linear scaling column.	
4	Table	There seem to be a number of errors in these formulae. Please check carefully.	Prof Ted Shepherd	These have now been corrected	
6	What is bias correction?	General comment: this is a very brief guidance document but nevertheless very useful.	Prof Dr Christoph Schär	The guidance is intended for those who do not know what bias-correction is but need to understand enough to know what may be required to undertake such a task and the potential pitfalls. There is a lot of published material providing much more detail and the user is signposted to it in the "Where can I get further information" section	
7	What is bias correction?	The methods recommended are not plain bias correction methods, but in general contain some implicit downscaling (certainly for QM and TPQM methods). This is relevant as the detailed physio-geographic properties of the target site are not represented in the climate models. This may include land-use, topography with shadowing and wind-channeling effects, nearby lakes or coastlines, etc.	Prof Dr Christoph Schär	We do not recommend any of the methods but introduce the user to the commonly used methods. Indeed QM and TPQM can be used for applications where there may be some implicit downscaling. The reviewer's additional information is now included in the text	
8	What is bias correction?	Following on from my previous comment, the differences between a model and a station record are not entirely due to model biases, but include small-scale climatic variations. A perfect model (at grid resolution) would exhibit differences in climate relative to single-station observations. As the effective model resolution is smaller than the nominal grid-spacing (by roughly a factor 5), this point is relevant event for the highest-resolution model provided (2 km grid spacing).	Prof Dr Christoph Schär	See comment 7	
9	What do I need to be aware of?	This list misses several important points: First of all, the mentioned bias correction methods merely correct the statistical distribution of a target variable (e.g. daily temperatures), but will not correct the statistical properties of time series (i.e. biases in autocorrelation are not actively corrected).	Prof Dr Christoph Schär	Agreed. An additional paragraph has been included to point readers at other types of bias-correction methods	
10	What do I need to be aware of?	Second, it should be stressed that the methods require sufficient observational data to characterize the reference climatology. In practice this is at least 10 years of data (better 30 years) in order to cover some of the decadal variations.	Prof Dr Christoph Schär	Agreed. This is now reflected in the text	
11	What do I need to be aware of?	Third the most common application of these methods use station data as reference. It is important to stress that the method presented are not suited in a multi-site context, as the temporal correlation between neighboring stations does not enter the method.	Prof Dr Christoph Schär	Agreed. An additional paragraph has been included to point readers at other types of bias-correction methods	
12	What do I need to be aware of?	It might be good to add that despite this limitation (the criticised assumptions), studies find it is much better to do some kind of bias correction / downscaling, rather than using raw model output.	Prof Dr Christoph Schär	This is expressed in line 22 but can be emphasised further. This is now reflected in the text	
13	What bias correction methods are there?	I agree that the list considered is not (and should not be) exhaustive. Nevertheless I recommend that you list some entry points into the literature that provides solutions for some of the most common problems. For instance, there is a large amount of literature on multi-site downscaling methods, a key challenge in many hydrological applications. There are a number of methodologies including both deterministic (e.g. Harpham and Wilby, 2005, DOI: 10.1016/j.jhydrol.2005.02.020) and stochastic procedures (e.g. Brissette et al. 2007, DOI: 10.1016/j.jhydrol.2007.06.035). Similarly, there is literature extending the QM method, so that it can be applied with short observational data records (Rajczak et al. 2016, DOI: 10.1002/joc.4417).	Prof Dr Christoph Schär	Thank you for the suggestions. These have now been added see response to comment 11	
14	What about downscaling?	As indicated in a previous comment, the more sophisticated methods contain some implicit downscaling. In addition, as this is a guidance document, I recommend, to stress that even the high-resolution model should undergo some bias correction.	Prof Dr Christoph Schär	Agreed that users should consider bias correction even for the high resolution models. 2.2km has also been added to the text	
15	Table	The table is useful, but some of the entries require revision. For QM/Cons it is stated that "Assumes future distribution is same as historical". That's not correct, it only assumes that the correction increments are the same as in the current climate, the distribution may change.	Prof Dr Christoph Schär	The table has now been corrected	

16	Table	For TPQM/Pros it is stated that "advantages using the QM method but also preserving climate change signal". I don't think that this is generally an advantage. If warm temperatures are overestimated at a station (e.g. due to the proximity of the coast), there is a good chance that also the climate-change signal is overestimated for the same reasons (i.e. QM might be the better choice)	Prof Dr Christoph Schär	The table has now been corrected	
17	Table	For "Linear Scaling/When to use it" you state "unsuitable for extreme events such as floods". This also applies to "Variance scaling"	Prof Dr Christoph Schär	Agreed. The table now reflects this.	
18	Table	For "QM/When" you state "use with caution for future climate studies". This is a general recommendation also applicable to TPQM (the basic assumptions are similar as with QM) and the other methods.	Prof Dr Christoph Schär	Agreed that this applies to all bias-correction methods and is in fact stated explicitly throughout the document. This has now been removed	
19	Table	For "TPQM/When" you state "for multi-variate studies". I don't think that this is the case for standard application of the methodology.	Prof Dr Christoph Schär	Agreed. In fact, the statement contrasts with the row above. Statement has been removed	

UKCP18 Science Overview report

Id	Section	Comment	Reviewer initials	MO response	Post Moderation Meeting Met Office comments
1	Summary	Can something related to the relative role/size of anthropogenic warming be added to this bullet point?	Dr Claudia Tebaldi	This deliberately aligns with the state of the climate report, which does not include attribution. However, earlier in the summary I have now mentioned attribution. Should I add some attribution refs?	Various comments added on attributution within the report, including the early section. I think these address this point.
2	Summary	Same as above: could you say something about the role of anthropogenic forcings in these changes?	Dr Claudia Tebaldi	As above, I want to keep this aligned with the state of the climate report but would be happy to added attribution references – would you be ok with doing it in the later section rather than key messages, as this won't be a key message from UKCP18 because it doesn't include attribuition.	as above
9	1 Introduction	"Variations in rainfall being more pronounced than variations in temperature" seems a bit of a fuzzy statement. According to what metric? Plus, the statement does not mean much divorced from any evaluation of the impacts of the changes, anyway.	Dr Claudia Tebaldi	Sentence deleted	
10	1 Introduction	This paragraph is a bit facile in mentioning a big issue, in my opinion, which is the need of somehow reconciling, or simply make sense of, some very spurious results across ensembles and strands. How are users supposed to do what is required of them? How are they going to go beyond the ranges provided? I'm not sure I would be comfortable with this paragraph as it is. I am tempted to suggest that the paragraph be cut after line 4, and a reference to guidance documents that will treat this issue extensively be made, and leave it at that.	Dr Claudia Tebaldi	Feedback during development has suggested some users, especially those more sceptical of models still need to have this stated. Further, one of the extra reviewers (DS) also stressed that we must highlight how difficult it can be to use climate information. There is growing literature suggesting the projections and the limitation should both factor into decision making. I have added some extra text at the end of the paragraph and a pointer to user guidance. Can you suggest some specific alterations to improve it?	Closed following discussion during moderation meeting on importance of highlighting difficulty of using climate model info in deicion making (DS reviewer) and user comments
25	2.5 How do users choose the appropriate UKCP18 projections?	Once again, this recommendation sounds a bit facile. Not sure how I (or anyone else) would do that.	Dr Claudia Tebaldi	But user guidance and follow on services will do this so important to state it. Have amended text to refer to user guidance.	
26	3.3 Projected Future Coastal Return Level Curves	This will be probably picked up in the Marine Report review, but it seems strange not to include uncertainty in present-day return levels, which I would expect to be large, especially if 1/1000 events are considered.	Dr Claudia Tebaldi	In the marine report we are adding a figure showing the present day uncertainty. However, combining this with other future uncertainties is a scientific challenge for the future.	Also this is reflected now in overview report figure
27	3.4 Projections of Changes in Wave Climate	The mention of CMIP5 models here driving simulations through wind output begs the question of how model validation is handled here. But more in general, how is model validation used in the Land strands too?	Dr Claudia Tebaldi	Details of the underpinning model validation is in the underpinning Land and Marine Science reports. The aim of the Science Overview is to draw on material and summarise material from the underpinning reports, and not to add new analysis.	
31	4 Notable differences between UKCP09 and UKCP18	I would show PDFs as well as CDFs as I find CDFs sometime less immediate to interpret, unless one focuses on specific percentiles.	Dr Claudia Tebaldi	To discuss as the PRP meeting, noting report is already long I would like to go with a majority view on this.	Disccsed at prp meeting and will make changes where agreed pdf works better. This is addition of figure showing the 1-year pdfs vs. 20-year pdfs. Discussion covered fact that it depends scientist on preference and also user preference.
32	5 Caveats and Limitations of UKCP18	A discussion of the different role of scenarios as the time horizon of the projections lengthens (or shortens) is needed here. The availability of a limited set of scenario is less of an issue for short term projections, and it would seem that the point should be made.	Dr Claudia Tebaldi	Text amended	
37	Summary	I think it is unwise to flag up such an extreme case in the summary.	Dr Ed Hawkins	Have removed the numbers from summary to avoid them being over used or taken out of context – good spot!	Whilst we have removed this example we note users will and should be able to generate their own examples through the user interface.
39	2 Climate scenarios over land	Figure 2.1 - I think the baseline choices are going to cause confusion, especially given the lengthy discussion of the Paris Agreement at the start. This first figure is not relative to pre-industrial, but could be. Perhaps add a second axis to highlight this? The thick orange line is not visible, and I don't understand the rationale for highlighting the STD version as it is not considered more plausible than the others? The figure also looks squashed. I think some consultation with graphical experts within the Met Office would be extremely beneficial for the entire report.	Dr Ed Hawkins	I think it is important to keep the baseline in figure 2.1 but I agree the idea of adding the pre-industrial is a great idea. We are also looking at figure quality throughout. Figure will be replaced in the final print version.	Detailed changes to figures agreed with EH post meeting, and reflecting views at the meetings. We have subsequently included as many of these as possible.
40	2.1 Observed climate change over the UK	Figure 2.2 - needs a scale on the figure, which also has too much unneccessary 'chart junk'.	Dr Ed Hawkins	Scale added to caption - figure format as in State of the UK Climate Report	See above. All captions also to be checked and refined. This is now completed. All captions sent to GH for comments and comments acted on.
41	General	The figures all need to be standalone - i.e. can be basically understood without a caption. Some of the colour and design choices for the figures in Section 2 are poor. Section 3 is better in this regard.	Dr Ed Hawkins	Figure colour choices to be reconsidered and captions to be extended in final versions	See above. All captions also to be checked and refined.
42	2.1 Observed climate change over the UK	Figure 2.3 doesn't even have a scale in the caption!	Dr Ed Hawkins	Caption extended. Increments are 10% in this plot.	Figure and caption further refined in final version to better match state of the UK climate publication.
43	2.1 Observed climate change over the UK	I thought that the observational timeseries were going to be extended further back before 1910 for temperature and rainfall for UKCP18?	Dr Ed Hawkins	See state of UK climate report – the overview here provides a summary of that peer reviewed publication. We have only provided a subset here for context of the projections.	

44	2.4 Plausible climate scenarios at other temperature levels	Figure 2.24-2.27 - the caption baseline looks wrong compared to Fig. 2.23. You also need a strong justification for why 1901-1920 can be considered as 'pre-industrial'.	Dr Ed Hawkins	Baseline here is different for the projections so the observations here are also presented on this baseline. Suggest we discuss wording for pre-industrial at the moderation meeting.	The 1901-1920 baseline is no longer used.
46	5 Caveats and Limitations of UKCP18	Is this statement on ECS true? I thought that HadGEM3.1 had an ECS of 6.2K which is outside the 'very likely' range of IPCC AR5? The ECS needs to be stated explicitly rather than just as a vague 'higher than' statement, and highlighted more prominently and earlier in the douments.	Dr Ed Hawkins	Sentence deleted. Text in bullet amended to be clearer on HadGEM3 and cmip5 contributions. The wider ECS issue to be covered at the moderation meeting.	ECS of HadGEM3.1 is not 6.2K. In the overview report we clearly show the warming of HadGEM3GC3.05 relative to the CMIP5 set for RCP8.5. We believe the range is consistent with interpretation of IPCC AR5. We believe focus on the warming for policy relevant scenario more appropriate for UKCP than focus on the warming for an abrupt forcing scenario.
47	Summary	I don't think a report as important as this can - in its summary - make promises not yet met or in any way trackable to deliver further information ("Later we will provide a further set….."). This report should only refer to itself, not concern itslef with work as yet not complete	Prof Kevin Horsburgh	I disagree – the CPM results have an agreed timetable, have been discussed with users and should be kept in discussions at this point. Also, just checked on run progress an many already now completed.	
49	Summary	H++'" I don't think the summary should use terminology that is no longer current. This report does not call the high end scenario H++, and not all readers have assimmilated the language of UKCP09. This bullet point could instead quantify the high end maximum and then just add that it does not differ significantly from the upper estimates that were given in UKCP09	Prof Kevin Horsburgh	Text amended	We have clarified the terminology of H++, which is well known to many users but not all so does need the new explanation.
50	Summary	However, there is some limited evidence…. As I said in my review for the marine section, this deliberately over-emphasises the selection of one non-downscaled member. The more detailed treatement on p63 is very good by the way. The caveat that COULD be included in the summary is that only a limited number of regional climate model members were used to force the storm surge model. Running a larger number may have produced a different spread of results. Bascially, future changes in storm surges cannot be ruled out although there is no evidence for it in the downscaled model runs	Prof Kevin Horsburgh	Have modified the text to follow the reviewers suggestion. I think this makes it clearer. Note the caveat applies to both surge and waves so gets an additional bullet at the end of the list rather an inclusion in each of the surge and wave bullets..	
54	3 Marine climate scenarios	To what extent is the use of an earlier HadCM version consistent with the use of GC3.05? This should be discussed in light of significant differences in physics (e.g. pertaining to the aerosol moduless used in HadCM3 and GC3.05).	Prof Kevin Horsburgh	Text amended	
56	3.1 Projections of Time-Mean Sea Level Change	Major comment: How do the estimates of ECS compare against the international literature? In particular, how does it compare against IPCC AR5 (Chapter 12)? In the AR5, a 66% probability has been stated of the ECS to be between 1.5 and 4.5K, a <1/10 chance to be below 1 K, and a <1/20 probability that the ECS is above 6K. The current estimate appears to be much narrower than that of the AR5. Can this be justified?	Prof Kevin Horsburgh	Text amended to improve clarity	The report shows the warming for RCP8.5 and transparanetly compares CMIP5 and GC3.05. This is more appropriate than a discussion in ECS space because it is the warming for the policy relevant scenario used.
65	General	As when I reviewed the other documents I have trouble with how LS2 is treated. Here in the summary it is given short shrift compared with the pdf approach and just slipped in. It is the really new part of this UKCP and its presence should be trumpeted! The term used for them at present if global realisations, I think. Perhaps this terminology should be considered. In any case its meaning should be spelt out. Ecery time it is discussed the fact that the relationships between variables and in time are retainehould be added to the list of pluses.	Prof Sir Brian Hoskins	Extra text added to summary – later in this report we highlight how the plausible ensemble members can be used by following the examination of 2 examples.	Terminology modified to global projections (LS2) and used throughout for clarity.
66	General	It must be clear that LS2 should not be treated in a probabilistic way as if it was another LS1. It could be good to refer to them as a set rather than an ensemble. Examples could be given of geographical maps of means for particular realisations that are the lowest, median and highest in some metric such as UK temperature change.	Prof Sir Brian Hoskins	This is now stated explicitly here and discussed again as we go through the report.	
67	General	as before I do not think that the issue of the climate sensiitvity of HADGEM3 is handled sufficitly upfront. For this reason, the set of global realisations has two subsets, one with 15 runs of HADGEM3 and the other with 13 CMIP5 runs. It should be clear that the set of runs is 15 +13. Then the 12km set is driven by most of the 15.	Prof Sir Brian Hoskins	To be discussed at the moderation meeting. We have made some additions that makes this clearer but further discussion is needed.	ECS was discussed extensively at the meeting. We agreed an approach that focussed on the comparison between the PPE and CMIP5 in terms of RCP8,5 response with a brief disucssion of what this implies for ECS. However, it is not appropriate to state values for ECS for the PPE has this has not been quantified for the PPE models. Also to note that warming to 2100 is a more relevant metric than ECS for the LS2 strand presented.
69	1 Introduction	Some simulations suggest that RCP2.6 scenarios may be consistent also with this target.	Prof Sir Brian Hoskins	Agree – text amended	
70	1 Introduction	The significant differences between Strand 1 and CMIP5 on th one hand, and Strand 2 on the other hand, are a source of concern. One wonders whether assessing the uncertainty primarily bssed on Strand 2 and the associated RCM simulation is really the right decision in this situation. The systematic differences are even larger when looked at in other metrics (Fig.3.22)	Prof Sir Brian Hoskins	Text amended	
72	2.1 Observed climate change over the UK	Need to mention summer 2018 at time of writing	Prof Sir Brian Hoskins	We will be preparing some specific comms material on this once the summer is over.	
74	2.2 Probabilistic projections of future UK climate	Need to mention some aspects of climate probably not captured well by any current model	Prof Sir Brian Hoskins	Caveat added	
75	2.2 Probabilistic projections of future UK climate	Fig caption needs more explanation & detail. Why is the blue line not in the middle of the blue runs? Why no obse after 2003?	Prof Sir Brian Hoskins	Figure to be updated Text added to caption to explain baseline	
76	2.2 Probabilistic projections of future UK climate	…which may partially reflect that…	Prof Sir Brian Hoskins	Text amended	

77	2.3 Exploring realisations of future climate over land	Fig. 4.3b: You need to specify whether these are all-day or wet-day percentiles, i.e. whether percentiles re taken over all days or only days above some precipitation threshold. In addition, I think that an intercomparison of precipitation frequency (P>threshold) would be desirable, as this is one of the standard metrics for evaluating precipitation climate.	Prof Sir Brian Hoskins	Good idea – text added to end of paragraph	
78	2.3 Exploring realisations of future climate over land	This is an example of a place where the climate senisitvity issue should be properly handled	Prof Sir Brian Hoskins	I've not mentioned ECS in order to be consistent with the land science report but do mention HadGEM3 samples warmer outcomes whilst CMIP5 covers mid-range and lower warming outcomes. Further discussion on this at the moderation meeting when discussing the land science report	ECS response discussed above in comment 67.
79	2.3 Exploring realisations of future climate over land	Not ensembles for LS2, and should be 15+13	Prof Sir Brian Hoskins	Text amended	We have been clearer throughout following PRP discussion that the global projections consist of a set of GC3.05 projections and a set of CMIP5 projections.
80	2.3 Exploring realisations of future climate over land	all driven from the 15 which are on the warm end of the set	Prof Sir Brian Hoskins	Previously added text on global ensemble – note is added here on sampling warming end of response.	
83	2.3 Exploring realisations of future climate over land	This part of the Land Report is extremely well written, and exhaustive. I read the whole thing without finding anything significant to comment on, UNTIL I started subsection 3.4, which presents the results of the ensembles evaluation. My comment here is very general, I'm not sure anything can be done about it, but I'm very concerned. I'm concerned about the behavior of the GC3.05 model, in its STD version and perturbed versions, for the historical period. The cooling for the most part of the 20th century at odds with observations, and the large regional biases connected to it, present a great challenge for the communication of these products, I believe. I'm wondering if, given that this behavior is the same in the STD version, single forcing experiments could be used to support the explanation given (aerosols) for this behavior. I'm not suggesting to run additional experiments for the UKCP18 project. Certainly single forcing experiments must be underway with the model in question. Having studies written up to which the authors of the UKCP18 project can refer, to prove what seems at this stage speculative, at least in the way it is described in this document, could help support the argument. I don't think that would solve the problem, but it could help. I'm also aware that the behavior here described is common to other models with interactive aerosol treatment. So the community has to reckon with it, not only the UKCP18 project. But the use of these results for practical applications presents a special challenge, and, stating the obvious, I want to underline the need for careful communication/justifications around this issue. Also the fact that the GC3.05 PPE has such a narrow spread, and a large portion of it (60%) comes from the perturbation of the concentration trajectories rather than the physics parameters perturbation is of great concern to me. I cannot evaluate if this latter result should suggest a failure of the PPE approach, or if the nature of the model is so rigid not to allow larger spread on the basis of the parameters chosen, and their values. But the small spread, together with the non overlapping quality of the two ensembles (GC3.05 and CMIP5) and their position with respect to the Strand1 envelope (Figure 3.19) will make the whole think difficult to digest by the users, I would think. I hope the communication of these products will be up to the challenge.	Prof Sir Brian Hoskins	Updated plot with better scale will be provided	In the guidance reports we have attempted to improve clarity on the different strands of evidence - and using them together. We've also been clearer on the pros and cons and discuss this extensively this most extensively in the land and marine science reports. We believe the use of GC3.05 and CMIP5 projections in LS2 provides a benefit over using either example along in informing decisions or communicating the climate science results.
85	2.3 Exploring realisations of future climate over land	This sub-section uses the probablistic intepretation of LS2 that I think is misleading	Prof Sir Brian Hoskins	It is not intended to do so – will try to change as I go through the section where it is unclear.	Further caveats added not to treat LS2 as a probability distribution - they are example futures.
86	2.3 Exploring realisations of future climate over land	CNRM-CM5 would seem to also simulate too strong an amplitude. The significance of it compared to the significnce of that of CCSM4 and CanESM2 cannot be assessed on the basis of the text, here, but it seems to be at least as significant as the weakness of MRI-CGCM3	Prof Sir Brian Hoskins	Text amended to be clear it is one realisation	
87	2.3 Exploring realisations of future climate over land	Have any real advantages been shown yet?	Prof Sir Brian Hoskins	Yes – see text in land report. However, this is not easy to demonstrate for the change signal.	
90	2.4 Plausible climate scenarios at other temperature levels	On 'near surface wind speed' need to say if this is average speeds or gusts.	Prof Sir Brian Hoskins	It is time mean	
92	2.4 Plausible climate scenarios at other temperature levels	mention mitigation and Paris	Prof Sir Brian Hoskins	Added on previous page.	
94	2.5 How do users choose the appropriate UKCP18 projections?	more needed in advantages of LS2 and 3 approaches	Prof Sir Brian Hoskins	The advantages of LS2 and LS3 are discussed in much more detail later in this section see para beginning "The plausible global and regional model projections provide flexible datasets..."	Text t further modified to also highlight the advantage of the LS2 and LS3 strands as not making the assumptions needed to construct the pdfs. This is also important to consider in response to review DS.
96	2.5 How do users choose the appropriate UKCP18 projections?	probably worth mentioning that the 2C and 4C approaches may remove most biases in T	Prof Sir Brian Hoskins	I think we should not bring this in with the probabilistic scenarios description here but I have added a new sentence in the synthetic scenario section to deal with this.	
97	2.5 How do users choose the appropriate UKCP18 projections?	it seems to get a bit repetitive at this point	Prof Sir Brian Hoskins	Yes, but important to emphasize this for users based on feedback from user groups and demo projects.	
98	2.5 How do users choose the appropriate UKCP18 projections?	..consistent iformation..	Prof Sir Brian Hoskins	Text amended	

101	Appendix 1	The document is long enough, so probably put on web	Prof Sir Brian Hoskins	To be discussed at the Moderation Meeting. Document is intended to be electronic but we take the point this may be better on the web.	
102	Summary	This makes it sound as if we are not already experiencing climate change! I write this during the heat wave in late July which has been marked by numerous media stories with respected scientists saying that this is the face of climate change. The validity of the statement made here depends on what one means by "dominate", and on the baseline chosen (and on the variable). It could be framed differently, and in any case should not suggest that climate change is not only something for the distant future.	Prof Ted Shepherd	Text added to end of previous paragraph indicates that climate change occurring. In this paragraph added mention of baseline to clarify what is meant.	This addition of mention of attribution has aided clarity for the reader.
103	1 Introduction	Same as previous comment	Prof Ted Shepherd	Dealt with previously. In this section I have made similar modification earlier in the section to highlight the obs changes and link to climate change. It is a very good point.	
104	1 Introduction	Isn't this statement more or less trivial?	Prof Ted Shepherd	Not for some users – I have encountered many for whom this still needs stating. Based on this I want to leave this in place.	
105	1 Introduction	You might wish to mention that the socio-economic assumptions behind the RCPs are reflected in SSPs which should be combined with the relevant RCP in any impact assessment (O'Neill et al. 2016 GMD doi:10.5194/gmd-9-3461-2016)	Prof Ted Shepherd	The O'Neill paper refers to the new AR6 scenarios NOT the AR5 RCPs, which are somewhat different and use older IAM runs to generate them. I don't think this reference would be appropriate. I've instead added a reference to the AR5 RCPs.	
106	2 Climate scenarios over land	I worry that Figure 2.1 may easily be used out of context. It immediately begs the question of how to interpret the rapid warming in the GC3.05PPE simulations. According to UKCP's own probabilistic projections, these plausible futures should be considered as extremely unlikely, so it may be asked why UKCP18 uses them so extensively. A box on the uncertainty in climate sensitivity could be very useful at this point in the report. For example, it could be noted that there appears to be no particular relationship between the uncertainty in climate sensitivity and that in the regional aspects of change for a given GWL, and thus that the advantages of GC3.05PPE in terms of storm tracks, etc., can be exploited without having to regard its climate sensitivity as a realistic prediction, by conditioning the results on a given GWL (as is done in the report).	Prof Ted Shepherd	After discussion with Prof Sir Brian Hoskins and Defra we feel the UKCP18 report is not the place to enter into the complex discussion of what HadGEM3 ECS value is. It is increasingly clear from the literature that this depends on how you measure it, and further we use a PPE for which we don't have the step experiments to reliably calculate ECS using the "IPCC" method. However, the point that the ECS is not an especially good indicator of regional response does need to be brought out more and I have amended the text to do so. But we do also need to show where the PPE sits relative to CMIP5 and if we are not using ECS what should we use? Further discussion on this to be covered at the PRP meeting.	Also see response to point 67. Note we have also moved Fig 2.1 later in the report so it fits in the with the main discussion of LS2.
107	2.1 Observed climate change over the UK	Comparing to a baseline of 1961-1990 looks like cherry-picking, once one looks at Figure 2.2. It is clear from Figure 2.2 that using this baseline maximizes the global warming effect, and that using an earlier baseline (e.g. 1931-1960) would give less dramatic numbers. The MO could be justifiably criticized for this. I suggest revisiting this choice of baseline.	Prof Ted Shepherd	The state of the uk climate report has been peer reviewed for journal submission and is now published. This statement comes from the published report and does mention two baselines. It also retains continuity with UKCP09 and a commonly used baseline.	Further discussion on baseline took place after PRP. This is reflected in the the land science report. The reader is pointed there for more info. We hope the impact of baseline choice is now clearer across the report and note users will be able to select their own baseline.
108	2.1 Observed climate change over the UK	This discussion of observed precipitation changes gives lots of numbers but no interpretation. For example, how unlikely is it to have two winters within a decade exceed 150% of the mean over a 30-year reference period? Without such context, what is the reader to make of this information? Also, since the climate projections in UKCP18 are for wetter winters and dryer summers, I would have expected some discussion of this given the apparent observed trend towards wetter summers.	Prof Ted Shepherd	Taken from and consistent with the state of the uk climate report. It would not be appropriate to go beyond the source material in the summary overview report.	
109	2.1 Observed climate change over the UK	It seems very weird to construct a continuous time series consisting of seasonal changes expressed in percentages, because the winter changes have a very different physical meaning than the summer changes. Suggest splitting the precipitation time series into winter and summer (especially given the context of the future projections).	Prof Ted Shepherd	See earlier comment – following the state of UK climate report for consistency.	
110	2.1 Observed climate change over the UK	Again, these observed changes in sunshine hours should be placed in the context of the projected changes. Moreover the most dramatic values are relative to the 1961-1990 baseline, which as noted earlier seems to be an anomalous period, climatologically speaking.	Prof Ted Shepherd	The observations are summarised for context. Future work will look at how events from the past might manifest in the future. Some of this will be done as part of package of launch material.	
111	2.2 Probabilistic projections of future UK climate	Whose understanding is "our understanding"? Sometimes this phraseology refers to the scientific community in general. But here I think it only refers to the architects of the UKCP18 probabilistic projections. Since it is already acknowledged that these probabilities are inherently subjective, you should not slide into terminology that somehow sounds more objective.	Prof Ted Shepherd	Fair point – text amended	
113	2.2 Probabilistic projections of future UK climate	The median line appears to be misplaced in the panel for winter precip, RCP2.6	Prof Ted Shepherd	Figure will be updated	
114	2.3 Exploring realisations of future climate over land	Do these bivariate distributions shown in Fig. 2.12 build in correlations between temperature and precipitation anomalies associated with interannual variability?	Prof Ted Shepherd	Yes it includes correlations between T and ppn anomolies associated with nat variability.	
115	2.4 Plausible climate scenarios at other temperature levels	Why has the baseline period changed compared to that used in earlier figures?	Prof Ted Shepherd	Figure replaced	
116	4 Notable differences between UKCP09 and UKCP18	You say that differences between UKCP09 and UKCP18 are due to tangible improvements, but can you actually trace the differences to changes that are demonstrably improvements?	Prof Ted Shepherd	There is some discussion of this in the land science report but I've amended the text to be less bullish.	
545	2.2. RCP2.6 time series	This situation is rather unsatisfactory. It acknowledges the need for spatial temporal scenarios but does not engage with the number of scenarios that are required for risk analysis (many more than have actually been provided) or with how users should choose scenarios from this rather odd bimodal ensemble.	Prof Jim Hall	This has been discussed at PRP on a number of occasions – the design philosophy is to provide a small-ish ensemble of the latest climate models rather than a larger ensemble of an older model. Users can consider how to use the realisations alongside the pdfs from LS1. Comment added to clarify this and point to the user guidance.	
548	2.3 Exploring realisations of future climate over land	Agreed that insights such as this are a benefit of the plausible realisations approach. But this also highlights the limitation of just having a very selective sample of the multi-dimensional variability.	Prof Jim Hall	I agree – comment added.	
549	2.3 Exploring realisations of future climate over land	Suggesting that looking at extreme values from only one realisation of natural variability in a transient simulation is potentially misleading	Prof Jim Hall	Text amended earlier in response to comment 548	

550	2.3 Exploring realisations of future climate over land	This is clearly a non-stationary series. No indication is provided here that the user should be using a non-stationary extreme value distribution	Prof Jim Hall	Good point – text added	
553	2.5 How do users choose the appropriate UKCP18 projections?	This seems to be rather dismissive of bias correction, when to this point no evidence has been provided of model biases. I note for example that precipitation is presented in terms of percentage changes rather than absolute values. What are the biases?	Prof Jim Hall	Sentence added to point to land science report to find out more about the biases.	Also note there is a specific fact sheet on bias correction as part of UKCP18.
554	2.5 How do users choose the appropriate UKCP18 projections?	This passage neglect the fact that many users used the WG for probabistic risk analysis. There are versions of PRA that users will now struggle to do with the UKCP18 products.	Prof Jim Hall	To be discussed at the Moderation Meeting. This is a point that does need addressing further.	Text around the weather generator has been revised.
556	Summary	As UKCP18 is also about observations of the past climate I think that there could be a short paragraph at this point with a few sentences stating stg on that, e.g. "… another decade of data … inclusion of new types of data?, … showing continued changes ...". Given such a paragraph the two first bullets in the list below could be moved up here.	Dr Erik Kjellström	I don't want the summary paras to become too long but have added a sentence highlighting the obs at the end of this paragraph.	
557	Summary	Isn't there a risk of confusion with the different wordings used in the different parts of the report. Aren't all estimates, realisations and projections used in UKCP18 "plausible"? I'm thinking that you could remove "plausible" here (and at line 11)	Dr Erik Kjellström	These are terms we have evolved with the user community – the idea of plausible realisations is distinct from pdfs. Therefore I would rather not change this. It is important to support the user more with this type of model output because they were not provided last time.	Terminology modified for global and regional projections to be clearer. Readers can look at the evaluation section of the land and marine reports (comparison with observations) when considering what plausible means.
558	Summary	… some … processes, … such as those …. Are there others? I would guess that this model is much improving the representation of wind, isn't this stg to mention here?	Dr Erik Kjellström	This needs to await the evaluation of the cpm. Also still need to maintain balance of readability with detail – this is the summary section of an overview report	To be further investigated in context of CPM.
559	Summary	Here it says "the change in future storm surge". Would be better with "possible changes in …". See also comment above on line 9	Dr Erik Kjellström	Text amended as suggested	
560	Summary	Shouldn't it be "projections for the UK"? There is some CMIP5 data in this isn't there?	Dr Erik Kjellström	Text amended as suggested	
561	Summary	Unclear if this alludes to the UK or somewhere else (global mean data?).	Dr Erik Kjellström	Added clarification.	
562	Summary	Here it says "some … some … some ". Does this mean that you can rule out also decade-long periods with cold periods, etc. Or, is that included in the term "some"?	Dr Erik Kjellström	This can be examined by users in the future when they are making decisions. We do not have this result at present. Happy to look in longer term.	Will also be interesting to compare the different UKCP18 components for this.
563	Summary	Here it says that UKCP18 differs from UKCP09. In what way? I think this would be valuable information for stakeholders who have already been working with UKCP09 to have upfront already here in the summary!	Dr Erik Kjellström	I agree – but it is difficult to summarise. For that reason we have a section on this later in the overview report and the land and marine science reports cover it. Should we link these bullets to sections of the report where the material is discussed in more detail? I would not be comfortable adding too much detail here as it is more complex than a bullet can cover.	We have continued with a distinct section on comparing UKCP09 and UKCP18.
564	Summary	Here it is stated "by 2100" and at line 21 it is rather 2080-2099. Why are you using different time frames?	Dr Erik Kjellström	Different products in UKCP18 allow use of different baselines and future time periods. This is reflected here. However, I have changed the text to refer to the end of the 21st century here. As these are summary bullets I feel this makes it more accessible whilst still being correct since we say "around 1 to 4". More detail is found later in the overview report.	
567	Summary	What is H++	Dr Erik Kjellström	Text amended	
568	Summary	If there is a number associated with this upper estimate it could be given here!	Dr Erik Kjellström	Text amended	
569	Summary	Could you be more specific here as to whether changes means increases/decreases?	Dr Erik Kjellström	Text amended - changed "changes" to "increases"	
571	2 Climate scenarios over land	I think it should say "inclusion of simulated natural inter-annual variability"	Dr Erik Kjellström	Text amended	
572	2 Climate scenarios over land	When I read the figure I get the impression that the HadGEM3 model tends to sample the warmer end as stated in the text on line 11. But, to say that the CMIP models sample the cold end implies that you give more or less equal weight to the (much) larger CMIP5 ensemble and the one-model HadGEM ensemble. The small spread between the HadGEM ensemble members indicate that the ensemble does not sample a wide range of outcomes. Shouldn't the perturbed physics approach better sample the full range of the other CMIP models? It is a bit worrying that it only covers a limited part of the CMIP distribution.	Dr Erik Kjellström	Yes we agree – it would have been nicer if the spread of the HadGEM3 ensemble had covered a wider range of global mean outcomes. The overlap is better on the UK scale and for some variables such as ppn. Understanding these points are topics for future research. However, it is still useful for users to have good coverage of the warmer range of outcomes that are physically credible but not covered by CMIP5. This is consistent with IPCC AR5 inflating the variance of CMIP5 results. For users I have amended the text to state cmip5 samples the mid-range and colder end.	
573	2 Climate scenarios over land	Why "should lead to improvements"? Is the HadGEM3 model better than what was used in UKCP09? In that case you may consider removing "should". Or, are you referring to other GCMs that could have been used instead of HadGEM3? In that case it needs to be made clear why and to what extent these are not as good as HadGEM3?	Dr Erik Kjellström	Text amended – "should" removed to make it clearer that we are comparing with earlier models. However, the land science report should be consulted for further evidence of this. We don't do a direct comparison on the information entering the RCM domain but do look at the global model response.	
574	2.1 Observed climate change over the UK	Is it just "been warmer"? Or, would there be room for saying stg along the lines of "statistically signficant warmer"? "considerably warmer"?	Dr Erik Kjellström	We have taken a decision with the authors and reviewers of the state of the uk climate report to not include attribution statements there. It is important we stay aligned with this. However, extra statement and reference added on attribution here to help clarity.	
575	2.1 Observed climate change over the UK	Why focussing here on a single year? When the report is out we are already at the end of 2017 and soon there will be demand for information about 2018… I would suggest staying with discussing decades.	Dr Erik Kjellström	The state of the UK climate report gives an update on most recent year and then talks about context. We do the same here, as discussed in earlier PRP meetings.	
576	2.1 Observed climate change over the UK	Same comment as above - why 2017?	Dr Erik Kjellström	See earlier response – based on state of uk climate report which discusses previous year in context of longer term trend.	

577	2.1 Observed climate change over the UK	I think that this whole paragraph should include some statements linking the precipitation climate in the UK to the larger global picture (e.g. "consistent with the observed intensification of the hydrological cycle"). It may be that not all readers are familiar with the large-scale changes.	Dr Erik Kjellström	Taken from and consistent with the state of the uk climate report. It would not be appropriate to go beyond the source material in the summary overview report.	
578	2.1 Observed climate change over the UK	Is 150% in one year exceptional? How often are there single years with more than 150% of the long-term average...	Dr Erik Kjellström	Added to the text that these are the highest. Looking at the numbers there are only 5 winter seasons that exceed 140% when looking back as far as 1910.	
579	2.1 Observed climate change over the UK	Same comment as for precipitation. Would be good to link statements about local changes in sunshine and wind to the larger global picture.	Dr Erik Kjellström	See earlier response for temperature. This goes beyond state of UK climate report.	
581	2.2 Probabilistic projections of future UK climate	Why is this so? Shouldn't an emulator represent the plausible outcome in a similar way as a physical model?	Dr Erik Kjellström	Emulators are not perfect. I think there is a greater confidence when a physically based model is considered as we can examine its credibility more completely.	
582	2.2 Probabilistic projections of future UK climate	What does it mean here with "present day baseline"? Better to use "1981-2000" if this is what is referred to!	Dr Erik Kjellström	Text amended	
583	2.2 Probabilistic projections of future UK climate	To what extent can the SRES scenario be compared to the RCPs in this figure? The forcing is evidently different but the question is if there are also other differences due to different underlying information (model verions, number of realizations, etc.).	Dr Erik Kjellström	Caption updated to state only difference between results is from scenario not other parts of methodology. There is a separate section of the report on that.	
584	2.3 Exploring realisations of future climate over land	The legend should be changed so that it is made clear which are "model A" and "model B" respectively.	Dr Erik Kjellström	New version will be provided	
585	2.3 Exploring realisations of future climate over land	I don't follow here. From what I can see the figure doesn't display the "seasonal mean summer temperature". Also, it is better to be a bit more explanatory and avoid using "mode of the distribution" so that also a wider audience can understand what is being referred to.	Dr Erik Kjellström	Reference to spurious figure removed – the figure is not shown – and mention of “peak” added to more general reader.	
586	2.3 Exploring realisations of future climate over land	These figures are very difficult to read as the points/circles gets very pale and to some extent (particularly for model B) ends up more or less at the same place in the diagram. You could consider adding ellipses encompassing one standard deviation of the data.	Dr Erik Kjellström	Updated figure will be provided	
587	2.3 Exploring realisations of future climate over land	Here, a footnote explaining NAO, NAO+ and NAO- as well as their implication for the weather/climate over the UK would be in place for any reader not familiar with details of the climate system.	Dr Erik Kjellström	Good idea – to be added.	
588	2.3 Exploring realisations of future climate over land	Are there any more members of A and B that has been (or could be) looked upon?	Dr Erik Kjellström	Not sure I understand this – there are other members that could be looked at instead of A and B (26 others). There are other weather types that users could examine too. Happy to amend text if reviewer can clarify what is needed.	
590	2.4 Plausible climate scenarios at other temperature levels	Why not also for other intermediate scenarios (RCP4.5 and RCP6.0)?	Dr Erik Kjellström	Decision taken at Governance board after consultation with users – this could be added later. RCP2.6 and RCP8.5 bound the RCP range	
591	2.4 Plausible climate scenarios at other temperature levels	The spread is difficult to read from this figure. I would suggest adding extra bar-whisker-diagrams for the two ensembles to the right representing the last 30 (or 50) years.	Dr Erik Kjellström	Updated figure to be provided	
592	2.4 Plausible climate scenarios at other temperature levels	In this last part of the paragraph it would be in place to give the global mean temperature increase between preindustrial (along with a definition of what you mean by it) and 1981-2000.	Dr Erik Kjellström	Does proposed change in Figure 2.23 address this?	
593	2.4 Plausible climate scenarios at other temperature levels	Here (and in Figures 2.25-2.27) it says 1900-1921. Is this "preindustrial"?	Dr Erik Kjellström	Clarified to “late pre-industrial” in all captions	
594	3.1 Projections of Time-Mean Sea Level Change	It would be in place with a sentence or two in this paragraph saying stg about why there are so relatively large differences between different parts of the UK in terms of increasing sea level. Also, how does the geographical pattern compare to observations?	Dr Erik Kjellström	Text added	
595	3.1 Projections of Time-Mean Sea Level Change	Recently, an error has been detected at CNRM in the way how data from their GCM (CNRM-CM5) was provided for dynamical downscaling. In this process SST (and sea-ice) has been taken from one ensemble member while atmospheric model levels has been taken from another one. This has not influenced the climatology (long-term averages) that looks reasonable in these simulations. However, all RCM simulations driven by CNRM-CM5 (including RCA4) show inconsistencies for time series so if there is a need for looking at day-to-day variability there may be problems. See further the documentation from CNRM at http://www.umr-cnrm.fr/cmip5/IMG/pdf/communication-issue-files_cnrm-cm5_historical_6hlev_en.pdf	Dr Erik Kjellström	Consulted with Marine report author and carry over changes to here. To be done once marine report have dealt with this issue. Marine report authors state "I've had a look at the link Erik sent. The problem is confined to the historical simulations so it has no implications for our projections of change (or rather lack of significant change), which are all based on the scenario runs (not the historical runs). The only part of UKCP18 marine which could be affected is the evaluation. The evaluation of simulated extreme surges against tide gauge data is based on the historical runs. All of the 5 Euro-Cordex simulations that we used evaluated adequately by this metric, and all 5 had a (slightly) smaller bias than the global-driven simulation (GFDL-ESM2M). So I think it's unlikely that correction of the bug (which in any case is not possible) would result in us excluding the CNRM-CM5 Euro-Cordex downscaled simulation from our set. Even if we did exclude it, it would not change the message for users; it would just be one less model. The other strand of evaluation, evaluation of the storm track metrics, is based on the global runs so that would be unaffected."	Marine report authors confirm no further change needed here to the overview report.
596	4 Notable differences between UKCP09 and UKCP18	may perform better feels vague. I think it should be possible to say either that it performs better or not.	Dr Erik Kjellström	After discussion we have decided to delete this sentence.	
597	5 Caveats and Limitations of UKCP18	I would rather express it as "introduces yet another level of uncertainty". There may be occasions when RCMs reduce the spread in the GCMs for a good physical reason thereby reducing also the uncertainty (e.g. Sorland et al., 2018, http://iopscience.iop.org/article/10.1088/1748-9326/aacc77/pdf)	Dr Erik Kjellström	Text amended	
598	5 Caveats and Limitations of UKCP18	I guess it should say "sea level and storm surges" as the reports do not look into other ocean issues (like sea surface temperatures, mixing, salinity, …)	Dr Erik Kjellström	Text amended	

Caveats and limitations

ID	Section	Comment	Reviewer initials	MO response	Post Moderation Meeting Met Office comments
14	General	There is a general problem with the approach taken in these guidance documents. The problem is also found in the land projections report and possibly others. The problem is that they are written as if they are trying to sell a product. Such an approach is fine in many circumstances, particularly for businesses selling a product in a competitive market, but the UK climate projections are essentially a government funded monopoly. That too may be fine because we might look to the Hadley Centre to provide impartial advice which reflects not only their own opinion and approaches but also those found elsewhere in the scientific community. The problem is when these too things come together. That these documents are written so strongly to sell the product has driven out their ability to reflect impartially on the pros and cons of the approach they take. This is a failing. I shall refer to this aspect as "sell" or "sales".	Dr David Stainforth	We take on this reviewer's critique of the way in which the Met Office models have been presented. We have responded to comments 14-37 as follows: (1) refraining from using adjectives that the reviewer has indicated as being ill-founded (2) focussing on providing caveats and limitations of the UKCP18 projections and not guidance (as suggested in comment 3) (3) including a reference to the critique of the UKCP09 probabilistic methodology - this is in the How to Use Land Projections guidance	
15	2 Keep in mind the caveats and limitations	Sell. The use of the term “latest scientific understanding” is misleading. It seems designed to inspire confidence and reliability yet the cutting edge of scientific research is almost always an area of dispute and disagreement. The use of this term in this way leads to a misrepresentative portrayal of UKCP18 and contributes to the widespread inability amongst the media and the public to understand that some aspects of climate change are reliably known and others are open to debate. It thereby undermines societal debate and consequently undermines the ability to take action to tackle climate change by creating a situation in which a product like ukcp18 is interpreted as of similar reliability to the knowledge that climate change is a serious threat to global society. I see this aspect of the documents as seriously detrimental to the aims of UK climate change policy.	Dr David Stainforth	This sentence has been removed such that the document focusses only on providing the caveats and limitations	
16	2 Keep in mind the caveats and limitations	Sell. The claim of “peer-reviewed by independent experts” is arguably an overstatement. Does the review panel have the option to reject the methodology outright as would be the case with a journal paper? Is there any option for any member of the panel to write a minority report (published along with the main reports) expressing dissent over the methodology or the value of the information presented? Some of the references are to papers due to be published in 2019 - is there any choice for the authors of those papers to respond to serious criticisms from anonymous reviewers in any way other than trying to get it published somewhere else and hope they don’t get the same reviewers? I have huge respect for the members of the review panel but does the review process really stand up to what would be considered international standards of independent peer review? I don’t know the answer because I don’t know what the process has been; that in itself highlights the problem with the process.	Dr David Stainforth	See response to comment 14. In addition, members of the peer review panel were provided in a project newsletter. We would appreciate guidance from the PRP at the Moderation Meeting on how to respond to this comment	
18	3 Climate projections are dependent on future greenhouse gases assumptions	Sell. It seems reasonable to use terms such as “plausible global realisations” and using “the latest climate models”, but peppering the documents with this type of phrasing gives a misleading impression of reliability and trust, even when discussing caveats and limitations. The impression is one of accepting concerns while trying to ensure the users don’t take them seriously.	Dr David Stainforth	These phrases been removed as suggested	
19	4 Estimated ranges for future climate are conditional	This paragraph is seriously misleading spin. I don’t think anyone has argued that the UKCP09 or UKCP18 teams deny the existence of assumptions or choices in their methodology. The “full acknowledgement” of this is simply spin to undermine the first sentence and to appear virtuous. By countering something that nobody has claimed it is easy to portray any critics as simply misunderstanding the situation. In fact the arguments put in discussions in the scientific literature is that many of these assumptions are known today to be ill-founded and the choices made are thought by some to be bad ones. Furthermore, the arguments are that different assumptions/choices would substantially (i.e. not marginally) change the information provided. Thus, it is not a question of “as the science evolves” but rather that under current understanding some believe the results should be very very different from those presented. It is sad to see that this draft of the document continues the approach taken in ukcp09 of not engaging with these debates or criticisms. These documents, perhaps unsurprisingly, reflect this lack of engagement.	Dr David Stainforth	The paragraph has been replaced with a version of the discussion now included in the science report on the land projections, i.e. it mentions the debates and criticisms of others in the scientific literature.	
21	4 Estimated ranges for future climate are conditional	Sell. This paragraph doesn’t really communicate anything apart from the idea that the information is useful. Users should of course choose an “appropriate decision framework” regardless of the conditionality of the information. What has this to do with the conditionality of the estimate ranges of future climate? How should users respond to such conditionality? How does it influence their “appropriate decision framework”? How can they map their “acceptable risk levels” onto the information GIVEN the conditionality that this section is meant to be describing? The result is that this section obfuscates rather than informs.	Dr David Stainforth	See response to comment 14. This section has now been revised to focus on the caveats and limitations. In addition, a new section has been provided in How to Use the Land Projections that put model projections into context of decision making which now separates the two discussions on (a) decision frameworks (b) conditionality	
22	4 Estimated ranges for future climate are conditional	The first paragraph provides a bit of a methodological summary and a statement that it relies on assumptions. That’s a good start but the second one undermines this by reverting to discussion of the brilliance of the Met Office model by comparison to other models and to earlier versions. This is not the point. They may be better but are they sufficiently good for the purpose to which they are applied here? The presentation here is pure sales talk.	Dr David Stainforth	See comment 14	

23	5 UKCP18 does not capture all possible future outcomes	This is fundamentally misleading. The probabilities don't represent strength of evidence based on the various aspects listed. They simply represent the distributions that pop out of the process chosen. They do not represent the relative strength of evidence in the sense in which it is conventionally understood. Furthermore the implication is made that they may not capture possible futures because of lack of current understanding. This fundamentally fails to represent the criticisms that have been made of the methods used, which are to do with complexity and current understanding of the limitations of the methodology. It is nothing to do with current limitations of understanding of climate processes. The failure to communicate this will no doubt lead users to conclude that it is widely acknowledged that ukcp18 is the best we can do today even while it may change in the future, but that is not something that is widely agreed. It is very disappointing to see the lack of engagement on these issues reflected even in the document on caveats and limitations.	Dr David Stainforth	See comment 14	
25	6 How to use the UKCP18 probabilistic projections and what they represent	Misleading sell. While it may be true that this is "the most comprehensive blending of key uncertainties and current modelling available", if it is ill-founded, as some have claimed, then the fact it is comprehensive is irrelevant. This seems to undermine what was acknowledged in the first sentence of section 4.	Dr David Stainforth	This sentence has been removed such that the document focusses only on providing the caveats and limitations	
28	6 How to use the UKCP18 probabilistic projections and what they represent	ECS is used here but I cannot see that it has been introduced/explained. The high ECS of GC3.05 is a crucial point and is just slipped in here. Please see my general comment.	Dr David Stainforth	See response to comment 26	
29	7 Climate model data contain biases	Is the difference in the value of CRF in GC3.1 quantitatively consistent with this being the/a major factor in the difference in its ECS?	Dr David Stainforth	This is reflected in the following sentence. However, we have removed "established" and modified the text such that this point is emphasised	
30	7 Climate model data contain biases	Sell. The implication is that it is simply a matter of choosing one of these four. In a document on caveats and limitations it is disappointing that there is no acknowledgement that none of them may be suitable.	Dr David Stainforth	Agreed. This is now reflected in the text	
32	8 Finer model resolution does not necessarily provide greater confidence	In a complex interacting nonlinear system how does this give greater confidence in a projection? (It of course has value in studying and understanding those processes but that it is a very different thing to making projections.)	Dr David Stainforth	This has now been removed from the text. See response to comment 14	
35	10 We have lower confidence in some products	Like "we cannot be sure" earlier, the highlighting of this aspect of the science being "at an earlier stage" will no doubt be taken as implying that the rest is robust. In fact it is all at an early stage of scientific understanding so disagreement (not just in the details) should be expected.	Dr David Stainforth	The earlier text has been modified to reflect that other methodologies may result in different results. The reader is also directed to the revised How to Use the Land Projections document. Stating this here about the 2.2km projections does not invalidate the previous statements.	
37	12 UK climate projections are likely to evolve	Again the implication is that the method is robust and the challenge is simply one of understanding the climate system, model quality and computational power. This substantially misrepresents the situation. The methodology itself has been questioned. There are a diverse range of views on how to approach this problem. It would be beneficial if DEFRA were to seek them out and encourage widespread discussion and debate on methodological approaches. Results could change with no change in understanding of the climate system, model quality or computer power.	Dr David Stainforth	We seek advice from the Peer Review Panel on how to respond to this critique other than what we have already included.	
40	6 How to use the UKCP18 probabilistic projections and what they represent	Shouldn't the likelihood of failing within a 10th percentile band be the same (i.e. 10%) irrespective of where that 10th percentile band lies in the overall distribution?	TGS	The reviewer is correct, the text as it stands is not helpful.	
41	4 Estimated ranges for future climate are conditional	It should perhaps be mentioned that the high climate sensitivity of the MO model is moot when one considers impacts of different GWLs, although there will still be structural model biases associated with the MO family of models.	Prof Ted Shepherd	Awaiting changes to Derived Projections report	An additional note has been included to reflect this.
42	10 We have lower confidence in some products	The time shifting assumption is really a very major limitation. There is growing evidence that for the CMIP5 models, the circulation response to a particular GWL is very different depending on whether that warming level is transient or stabilized, e.g. Ceppi et al. 2018 JClim DOI: 10.1175/JCLI-D-17-0323.1. Because this report will be consulted in a few years time, when that evidence base will likely have grown further, some acknowledgement of this limitation is needed here.	Prof Ted Shepherd	Awaiting changes to Derived Projections report	Text has now been included to reflect the major assumption

Derived Projections of Future Climate over the UK

ID	Section	Comment	Reviewer initials	MO response	Post Moderation Meeting Met Office comments
5	3.1. Changes to UK climate in 2°C and 4°C worlds from the HadGEM3.05 model	Figure 1: the colour scale needs changing so that the uncertainty is clearer and the values can actually be disti guished. I realise this may be a standard, but this makes the information impossible to read. Also, the two extreme ends (pink and green) are not appropriate.	Dr Ed Hawkins	The colour scales and data ranges originally used in this report are the same as used by land strand 2 so that there would be consistency between the maps of in each report. However for this temperature map they are quite washed out as the levels warming in the 2C time slice are much lower than those in RCP8.5 examined in LS2. We have changed the range and increments of the colour scale to better pick out the detail. These changes have also been made to the figures for 4C of global warming (figure6). In order to produce standardised graphics across reports we are working with land strand 2 to determine a consistent set of figures and land strand 2 will be presenting options in this regard are the review meeting. Once a standardised format has been determined all the map figures in this report will be updated. At present only those for temperature and precipitation have been changed.	

9	General	My major comment is on the way the results are shown. Repeating my major general comment on Land Strand, the results are not dscribed and discussed in a way that is consistent with the storyline nature of Strand 2 and the way that it is envisaged that users should approach it. Ensemble means/medians or lowest or highest 10 percentiles should not be considered or shown. The ensembles are not set up for this, and it is misleading. Instead, maps for the member that has the lowest, median or highest value in the UK average of the quantity in question should be shown and discussed. I will not repeat this comment every time it is relevant, as it applies to every figure with UK maps.	Prof Sir Brian Hoskins	The maps that we present show the 10th, 50h or 90th percentiles of the frequency distribution of the GC3.05-PPE calculated for each grid cell, giving an indication of the spatial pattern of the upper, central and lower portion of the frequency distribution. To demonstrate their usefulness we have added figures for the difference between the 90th and 10th percentiles to show the spread in the ensemble for a given variables at the two levels of global mean warming. Were we to use the UK mean of a variable to pick upper and lower ensemble member, the spread at any single location may be underestimated giving a misleading impression of the uncertainty within the ensemble. However, we have also included some new figures showing the cumulative frequency distribution for the UK mean of temperature and precip from which individual simulations are labelled so that they can be selected and used for a storyline approach as needed by users. Figures showing the CDFs of UK means where ensemble members are identified will be made available for all variables on the web site. Responses in LS2 about the appropriateness of the 10th 50th and 90th percentile maps are also relevant here.	We have chaned to showing "exemplar" simulations which have UK mean changes for a given variable nearest the 10th, 50th and 90th percentiels of a cumulative frequency distribution. These are then compliments with maps based on the grids cell level spread to highlight (where needed) the potential issues with exemplar approach. i.e. that they can give a misleading indication of the regional unceratainty.
10	Summary	Amplified by about how much?	Prof Sir Brian Hoskins	Reordered some of the text to make clear that in broad terms the changes seen in a 4C world are roughly double those in a 2C world.	Now reporting changes relative to present day so the near doubling is less apparent,
11	Summary	for the CMIP models	Prof Sir Brian Hoskins	I think this is phrasing issues rather than a scientific issue. As such we've amended the text here. To be clear here on the issue, as is we have elaborated on in the text, the reason that we only use HadGEM3 for the two warming level time slices is that we need the global mean temperature to have passed a given that level in question to produce the time slice. As not all of the CMIP5-13 pass 4C we cannot use them for the time slices. This means that there would be an inconsistency in the ensemble members used for 4C and the 2C time slice as well as for the RCP2.6 data, so we chose not to include CMIP5-13 in the time slices rather than introduce confusion for users.	We are now using the CMIP5-13 models which pass both 2 and 4C as well as the HadGEM3 simulations for the time slices.
15	Summary	Here, elsewhere and in particular in the Appendix title, I do not like the use of the word validation. It means to confirm the truth of, rather than to evaluate or test how well it works in some circumstance.	Prof Sir Brian Hoskins	The aim of the methodology used is to reproduce what the models would have simulated, so in that sense there is an argument that "validation" could be appropriate as model simulations are the "truth" we're looking to confirm, but we concede that the use of validation might over emphasize the strength and generality of the assessment. As such we've now changed the use throughout the document to discuss "evaluation" instead of "validation".	
16	1 Introduction	easterly probably better	Prof Sir Brian Hoskins	Surface wind data components from across UKCP18 as well as CMIP6 are provided as Eastward and Northward components (Westerly and Southerly) so we are keeping this this convention.	
17	2 Methodology	somewhere in this paragraph or near here there should be a short discussion of climate sensitivity as this was a major reason for including CMIP models in LS2	Prof Sir Brian Hoskins	We have added text before and after this to make clear both that the small range of ECS from the HadGEM3-PPE was a factor in the inclusion of CMIP5 data into LS2, and to stress that this is not an issue for the synthetic time slice data in this report. By presenting information at warming levels we remove the ECS uncertainty and instead place it into the consideration of timing of reaching these warming levels. In the discussions (section 4) we've added a table of values indicating the range of timing of passing 2C and 4C from CMIP5 and GC3.05-PPE to aid user application of the time slice data.	
20	3 Results	This is the section my major comment mostly refers to	Prof Sir Brian Hoskins	(see response to comment 11)	
25	UK climate changes at 2°C of global mean warming	The Fig 1 T scale does not enable any structure to be seen. Given that the largest values in any later Fig are, I think less than 6C, this should be to top of the scale. The same scale should be used in the other T Figs.	Prof Sir Brian Hoskins	Done	
32	Timing of reaching 2°C and 4°C of global mean temperature increase	It would be helpful to give more info (e.g. ranges of Years) in which 2C and 4C may be achieved under various RCPs	Prof Sir Brian Hoskins	The end of section 3.1 (results) now includes some text discussing the timing of reaching 2 and 4C and includes a table indicating the timing for2 and 4C for GC3.05-PPE, CMIP5 and CMIP5-13 for RCP2.6 and RCP8.5.	
36	General	It would be good to discuss briefly resolution and in particular the idea that higher res my show enhanced convective rainfall in summer	Prof Sir Brian Hoskins	Done. Added to discussions in section 4.	
47	1 Introduction	I don't think it is appropriate to invoke the PRP within the report itself. This is passing the buck. The PRP has been overseeing the process, but its options are limited. The MO must take responsilbiity for its own report.	Prof Ted Shepherd	This was intended not as a means of "passing the buck" but rather to give an accurate description of the decision process. We will remove reference to the PRP.	
48	2 Methodology	It is not correct to say that time shifting was done "at the request of the PRP". This is passing the buck. The PRP was asked to comment on various options, but its options were limited. The MO must take responsilbity for its own product.	Prof Ted Shepherd	Again, we'll remove any reference to the PRP from the text.	

49	2 Methodology	Time shifting clearly introduces errors: see e.g. Ceppi et al. 2018 JClim DOI: 10.1175/JCLI-D-17-0323.1. So the question is how large these errors are.To say they are smaller than other sources of error depends on things like ensemble size, so is not a very informative statement. A qualitative assessment of the potential kinds of error would be useful.	Prof Ted Shepherd	Our suggestion to add CMIP5 models to Strand 2 was motivated by a general expectation that including structural model diversity would broaden the range of realisations. This motivation related to regional *patterns* of change (for example, without CMIP5 models, Strand 2 would not include any realisations in which UK summer rainfall does not dry in the future), as well as global temperature outcomes. At the time of decision, we did have information from idealised experiments suggesting the possibility of some high-end global warming outcomes, but we didn't yet have the transient GC3.05-PPE simulations. Therefore, we don't think it would be appropriate to discuss the high warming earlier, in the *experimental design* part of section 3, as if it was the sole (or dominant) reason for adding the CMIP5-13 models. This is why we introduced the discussion of high warming in section 3.5, rather than earlier in section 3. We do, however, agree that the high warming result should be mentioned earlier in the report. We will address this by adding an Executive Summary at the front of the document, as suggested in comment #230. In this report, we are happy to include open and confident discussion that we can support with peer-reviewed evidence. However, we are constrained by the newness of the GC3.05 simulations and the project delivery timescale. We think it would be unwise to include "work in progress" information, due to the risk that some of this might later have to be retracted. In particular, we do not agree that the UKCP18 land science report is the right document in which to provide an extensive discussion of ECS. This is motivated by three reasons and follows discussion with the Met Office Chief Scientist on the issue of presenting ECS: • Work on understanding the ECS has not used GC3.05. Rather, it has used GC3.1 and UKESM1. • Aside from the standard model variant, GC3.05-PPE consists of members that each have different ECS values, but these values cannot be estimated from the scenario experiments available. • Work on understanding ECS in GC3.1 is not yet peer reviewed or published and is still evolving. Presenting a non-peer reviewed version risks being out of date and incomplete. The work will be published in due course and is relevant to multiple Met Office projects, including our CMIP6 contribution. However, we do agree with the need to address the high warming issue in the UKCP report and show how the PPE compares to CMIP5 models. We propose to do this by showing a comparison of the RCP8.5 warming response in the GC3.05 PPE with the IPCC AR5 simulations. This is a more appropriate comparison because it uses the scenarios actually applied in UKCP18. This marks a change from the draft report, in which the ECS estimate of 6.2K (for GC3.1) was included. We based	We added the following text, to give a sense of where ECS values for GC3.03-PPE may lie with respect to the current IPCC assessment: "In AR5, IPCC assessed ECS to have a likely range of 1.5-4.5ºC (Collins et al., 2013), and also judged thatthere is a small probability (of up to 10%) that ECS exceeds 6ºC. The levels of 21st century warming simulated for the RCP8.5 scenario suggest that most of the GC3.05-PPE members are likely to possess ECS values above 4.5ºC. ". At the Moderation Meeting, the PRP advised us not to add an Executive Summary, so we have not done this.
51	Mean climate	Just to be clear: you are detrending all variables, correct? How can we be sure that the detrended fields are physically self-consistent?	Prof Ted Shepherd	The temperature and precipitation monthly data are consistent due to their derivation. We have already highlighted that this might affect the consistency between the other variables in Section 4 of the document. Hence, we can't be sure but we have tried to be as consistent as possible within the time frame we had. The time mean value for each variable is physically consistent because it is an average across the same time period for all variables. In the internal variability calculation, the same time period is chosen for all variables and all variables are detrended using the global mean temperature response.	
54	Mean climate	To what extent is the use of an earlier HadCM version consistent with the use of GC3.05? This should be discussed in light of significant differences in physics (e.g. pertaining to the aerosol moduless used in HadCM3 and GC3.05).	Prof Ted Shepherd	This is discussed in section 5.1: "The HadCM3- and GC3.05-based PPEs were both developed from UK climate models. Their land surface components are structurally similar, however the two models contain many structural differences, including their representations of atmospheric dynamics, large-scale cloud and boundary layer processes (Walters et al., 2017). The overall ranges of response seen in Strands 1 and 2 for GMST (Fig.3.19), and patterns of response (Fig. 5.1), are therefore a consequence of combining three quasi-independent lines of modelling evidence, namely: HadCM3-based PPEs (Strand 1), GC3.05-PPE (Strand 2) and CMIP5 multi-model ensembles (Strands 1 and 2, via emissions- and concentration-driven experiments respectively)."	We have recalculated everything to show changes relative to the present day instead of pre-industrial.
62	3.1. Changes to UK climate in 2°C and 4°C worlds from the HadGEM3.05 model	The results at 4C seem largely just a doubled version of the results at 2C. Rather than just repeating the description, could differences be discussed? For example, if the extreme percentiles reflect a large component of internal variability, then one might not expect them to scale with GWL.	Prof Ted Shepherd	The design of the methodology means the internal variability with the two time slices is correlated. As such there is no scaling of internal variability with GWL.	
64	Daily temperature distributions at a global mean warming of 2°C and 4°C	The difference between daily and monthly extreme wintertime temperature changes makes physical sense (i.e. the coldest days warm the most), and this should be stated. Are there differences between daily and monthly extreme precipitation changes that also align with expectations from basic physics? If so, that should be noted.	Prof Ted Shepherd	We agree that some more physical understanding would be beneficial but there was no time or resource for this to happen before the review. This will be taken forward at a later stage as the work is written up for publication.	
65	RCP2.6 projections	The distinct downward kinks in both UKCP18 and CMIP5 in the early 1980s and again in the early 1990s in Fig 15 should be commented on, because they will attract attention.	Prof Ted Shepherd	The kinks in temperature in early 1980s and 1990s are due to the response to volcanic eruptions of El Chichon and Mt. Pinatubo respectively. These will be noted in the text when mentioning the figure.	
66	RCP2.6 projections	Given all the discussionearlier in this document about the near cancellation between winter wetting and summer drying over the UK, it makes no sense to present annual mean changes as the best summary of our state of knowledge concerning the effect of climate change on precipitation, as in Fig 16. Suggest showing winter and summer separately for precipitation.	Prof Ted Shepherd	We will modify this section to show the winter and summer time series separated.	
67	4 Discussion and limitations	The larger warming in summer than in winter is the opposite of what one expects from elementary physical considerations (such as those behind Arctic amplification), so perhaps some physical explanation would be appropriate?	Prof Ted Shepherd	We agree that some more physical understanding would be beneficial but there was no time or resource for this to happen before the review. This will be taken forward at a later stage as the work is written up for publication.	
68	4 Discussion and limitations	It is misleading to represent the drying in summer as "opposing" the wetting in winter. Each seasonal change will have its own climate impact; they do not cancel out!	Prof Ted Shepherd	A clumsy choice of words. We'll rephrase this.	
69	4 Discussion and limitations	It is misleading to emphasize the relative changes in precipitation; the absolute changes are probably more user-relevant.	Prof Ted Shepherd	We chose to show relative change for precipitations to be consistent with the LS2. We understand that there are advantages and disadvantages in showing either relative or absolute changes and are keen that users have access to both so that the can choose the most appropriate for the application, so data is all provided as absolute values. However, we will follow what ever is agreed for LS2 regarding this.	
70	Issues effecting only time-slices at 2°C and 4°C of global warming	The significant differences between Strand 1 and CMIP5 on th one hand, and Strand 2 on the other hand, are a source of concern. One wonders whether assessing the uncertainty primarily bssed on Strand 2 and the associated RCM simulation is really the right decision in this situation. The systematic differences are even larger when looked at in other metrics (Fig.3.22)	Prof Ted Shepherd	Good point we will add some commentary about this in the revised text.	we've added text in a few places about this.

71	Potential improvements	Surely that should be done within UKCP18?	Prof Ted Shepherd	We will add comparisons with the RCP2.6 CDFs for 2050 and 2090 seasonal temperatures for the UK mean.	CDFs have been added and the overview now carries material comparing the spread to that from LS1 and LS2. The resutls from the derived RCP2.6 are within the LS1 range.

How to use the UKCP18 land projections

ID	Section	Comment	Reviewer initials	MO response	Post Moderation Meeting Met Office comments
4	2 What should you be aware of before using the land projections?	General comment: the tone of this section is pretty much dominated by the process of scientifically generating these scenarios. One could also have chosen the perspective of the user instead. Indeed, this guidance document is a product description rather than a guidance advice for users	Prof Bart Van Den Hurk	The reviewer has a good point, with much of the document devoted to describing the differences between the land strands and the advantages/disadvantages of using them. This was the original scope for the document. We have taken this on board and will include a decision tree early on in the document to help users navigate the land projections (and the material in the guidance) for typical use-cases.	
9	2 What should you be aware of before using the land projections?	General comment: I find it remarkable that you don't mention bias correction (or sensitivity of impact assessment to the potential biases in the storyline projections) in this introductory set of remarks. You do imply that these model outputs can be used and compared to the real world straight away, which may lead to bad results when naively interpreted	Prof Bart Van Den Hurk	Biases are in the first paragraph of the section which refers to biases in climate modelling in general and also in sections 5.2 with specific warnings. The point is noted and biases are mentioned again under the global realisations, referring the user to the bias-correction guidance.	
15	4 How do you choose the appropriate UKCP18 projections?	General: I wonder whether this section has been seen by actual users. Wouldn't a number of past examnples from UKCP09 not be a better illustration of what kind of information can support which kind of application? The tone of this intro section is very generic, not very informative	Prof Bart Van Den Hurk	This section has now been replaced by a decision tree to help readers navigate the high-level decisions to arrive at the section in the guidance that they require. Examples are now placed upfront to illustrate their use .	
36	1 Why are there three strands of information?	Sell. “Comprehensive” sounds good but sidesteps the more important questions about reliability. The phrase “given our current knowledge” is misleading because it implies that this is the best information available whereas “current knowledge” might be used by others to conclude that a comprehensive assessment of this form is inappropriate or partially inappropriate. This is an example of changing the issue to an irrelevant one i.e. whether this is the most comprehensive rather than whether it is the best approach to seeking information to support adaptation planning. I agree it is the most comprehensive but I think parts of it are a bad approach to supporting adaptation planning.	Dr David Stainforth	The document has been revised to address the comments 36-39, 41-44, 46-48 by: 1. including a section on using climate model projections for decision making/adaptation planning to outline 2. removing adjectives such as "relatively", "comprehensive" and tighter usage of "robust" 3. including a reference to the debate about the assumptions underlying the UKCP09 probabilistic projections but removing text describing the probabilistic methodology i.e. "relative strength of evidence supporting different plausible outcomes for UK climate". Further revisions to the current text will also reflect any changes made in the land report	
37	1 Why are there three strands of information?	There is a tendency to judge the results in terms of improvements over what has gone before rather than in terms of relevance and value to decision makers. In this table the term “relatively comprehensive assessment of model uncertainties” is both accurate and misleadingly irrelevant. What is the implication of a “relatively” comprehensive assessment?	Dr David Stainforth	See response to comment 36	
39	2 What should you be aware of before using the land projections?	It would seem appropriate to point out that these are not just conditional on these things but that the methods and choices are contested.	Dr David Stainforth	See response to comment 36	
40	2 What should you be aware of before using the land projections?	What constitutes “reasonable variations”? The implication is small variation. I disagree. Surely there should be guidance on this. How can users know how to make this judgement?	Dr David Stainforth	See response to comment 8	
41	3 What is the relationship between the three sets of land projections?	I’m surprised to hear that anyone would ask for a product exploring “uncertainties relatively comprehensively”. What does that mean? How would one use it? I think this points to a need for DEFRA to be clear about user requirements in any future product. In UKCP09 and UKCP18 the user demands seem to have been very narrowly interpreted as what variables, resolution, means of presentation etc. are required. In the future it needs to cover reliability, robustness, interpretation etc.	Dr David Stainforth	See response to comment 36	
42	3 What is the relationship between the three sets of land projections?	I disagree. I do not think these are as robust as possible and I think there are dangers in using them in risk assessments. These dangers have not been addressed in the documentation so there is a risk they will be misused.	Dr David Stainforth	See response to comment 36	
43	4 How do you choose the appropriate UKCP18 projections?	Sell. How robust? This claim of robustness contradicts the information in the caveats and limitations.	Dr David Stainforth	See response to comment 36	
44	4 How do you choose the appropriate UKCP18 projections?	“subject to the caveats listed in this document”. There is no guidance as to how users should respond to those caveats and the way the document is written leads one to think they aren’t very important. This is a significant failing in the documentation.	Dr David Stainforth	See response to comment 36	
46	4.1 When should you use the probabilistic projections?	I fundamentally disagree with this statement.	Dr David Stainforth	The document has been revised to include a more specific discussion on the probabilistic projections are in Section 2. The text here has been modified to focus only on the caveats rather than explaining what they are. We leave the discussion about whether the reviewer agrees with our approach with the probabilistic projections to the discussion in Section 2.	
47	4.1 When should you use the probabilistic projections?	This phrasing implies that they are good estimates if not mathematically exact. The point under contention is not whether they are mathematically exact but whether they are even reasonably good estimates. The documentation fails to communicate this. The “for example” clause makes it worse by highlighting the case of influences which are not understood while the main criticisms do not relate to such unknown unknowns.	Dr David Stainforth	See response to comment 36	
48	4.1 When should you use the probabilistic projections?	Sell. If it doesn’t effectively do modelling uncertainty then its inclusion of the other uncertainties is irrelevant. The document needs to acknowledge the disagreements and debates in this field.	Dr David Stainforth	See response to comment 36	
49	4.3 When shoud you use the regional realisations?	How does downscaling increase the level of uncertainty? In what way? How should a user interpret this statement?	Dr David Stainforth	We added the following text, to give a sense of where ECS values for GC3.03-PPE may lie with respect to the current IPCC assessment: "In AR5, IPCC assessed ECS to have a likely range of 1.5-4.5ºC (Collins et al., 2013), and also judged thatthere is a small probability (of up to 10%) that ECS exceeds 6ºC. The levels of 21st century warming simulated for the RCP8.5 scenario suggest that most of the GC3.05-PPE members are likely to possess ECS values above 4.5ºC. ". At the Moderation Meeting, the PRP advised us not to add an Executive Summary, so we have not done this.	

50	4.3 When shoud you use the regional realisations?	Sell. “which provide a better representation of convective processes”. The land projections science report notes that there are provisos on the conclusions one can draw from this. It is interesting to note that in this document they are not mentioned.	Dr David Stainforth	The caveats are now more explicit in the text and reader is referred to Kendon et al (2017) as mentioned in the land projections science report	
92	4.1 When should you use the probabilistic projections?	If you're going to do a flood risk assessment you need time series inputs for a hydrological model. LS1 doesn't do this. Kay and Jones used transient model outputs i.e. LS2	Prof Jim Hall	The reference has been corrected and now references a paper with same authors and year where they compare different UKCP09 products using probabilistic projections, the weather generator and the transient RCMs.The paper suggests that more than one UKCP09 product should be assessed in an application - this is now included in the text.	
93	4.2 When should you use the global realisations?	But the global realisations don't provide enough realisations to do this.	Prof Jim Hall	The global projections provide data to explore a range of climate outcomes and enable users to drive their impacts models using time series information. We suggest that these should be treated as storylines rather than be used to assess relative likelihood. This is made clearer in this sentence.	