Thank you to both anonymous referees for their suggestions and discussion. In the following pages answer their questions and comments in italics.

Anonymous Referee #1

Review of 'Using the UKCP09 probabilistic scenarios to model the amplified impact of climate change on river basin sediment yield' by Coulthard et al.

This is an interesting and timely paper. It examines the issue of rainfall change under different potential climate scenarios and its effect on sediment transport and ultimately water quality. Such research is essential as using modelling approaches such as this is the only way that future climate and its effect on rainfall can be examined in a quantitative and statistically testable framework.

This is a well-written and easy to read paper that should be published. There are some minor comments below that the authors would do well to address before publication.

1. Section 4, 2nd para. The 50m grid scale is big for a hillslope and channel scale assessment such as this. It is fully recognised that the catchment is HUGE for such a study, and the run times are long, but can a 50m DEM provide the information/detail of hillslope and channel? Comment needs to be made about this.

Thank you for the very positive comments – for such a study this is a large catchment area . Therefore , choice of model resolution is a balance between spatial detail and run time. Halving the grid cell size unfortunately results in an increase in grid cell number (and thus calculations) to the power of 2. We also chose this resolution and location as we had used a slightly different model of the Swale before (Coulthard and Macklin, 2001). However, there is always a loss of detail for slope and channel features that will be concealed within the averaging effect of a larger grid cell. Whilst this may be important if making detailed interpretations about select parts of the drainage basin, when comparing relative changes in sediment yield (in this case between baseline and different climate scenarios) the effects of scale/resolution are less.

2. As a modeller I understand the great strengths of this study and the approaches used. However, it would really help to convince the field people and non-modellers if there were some field data to help back these findings. Can evidence be derived from high magnitude floods to help convince the sceptics? This is a suggestion only as this may be difficult to do.

Long term, high resolution measurements of bedload are non-existent in the UK making it very difficult to directly back up these results. However, examining the Palaeo flood record provides evidence of very large floods in the Northern England uplands (Rumsby, 2000) and across the UK (Macklin et al., 2005). These have been linked to past increases in flood magnitude and frequency associated with the Little Ice Age that the authors suggested may be an analogue for future increases in flood magnitude.

Anonymous Referee #2

This paper discusses the potential changes to sediment yield in a small river basin caused by climate change. It is a very important topic, and the method employed is according to the authors novel, which would merit publication in HESS. First of all, I would like to state that the paper is nicely written, and the introduction is excellent, it puts the paper in context. However, I have some remarks on the experiment setup as well as the interpretation of the results. I therefore recommend that the paper undergoes a major revision and is then resubmitted.

Scientific content The first major remark is the use of the UKCP09 generated series. I understand that a sample of the climate runs has to be made because of limitations to computational time, and I agree that the sample selection is reasonable. However, it is not clear to me how the precipitation series were put together for a specific run. The authors mention that 5 tiles were used that overlapped the area. Was the driving data then the average rainfall from these 5 tiles? As I understand, the weather generator is a stochastic tool, which means that the grid points are uncorrelated in space and time. This makes them not useful for creating realistic catchment averages of rainfall, and therefore also not useful in hydrological studies. Moreover, how do the points overlay the catchment? And how is the overlying 25 km grid, from which the change factors are calculated, related to the 5 km grid and the catchment? These points are not clear to me, and the effect they may have on the rainfall and runoff needs to be clarified.

We use the standard UKCP09 approach to produce precipitation series using the weather generator. We use five 5km by 5km tiles which overlap the catchment. The way the weather generator works is that statistics for each individual tile are averaged and the weather generator is then fitted as a point model to this 'averaged' set of statistics. Of course, this method only works in small catchments (less than 40 5km cells can be selected and they must be contiguous) as in larger catchments we would expect heterogeneity in rainfall and thus a multi-site weather generator would be needed (as suggested by the referee). Therefore, we use only a point rainfall estimate (which can be thought of as a lumped average over the catchment) as input to the hydrological model. We have added a sentence into the text to clarify this point "In the case of a small catchment (defined as up to 40 5km grid cells), statistics for each individual 5km grid cell are averaged and the weather generator is then fitted as a point model to this 'averaged' set of statistics."

To answer the second point of the referee, the way the overlying 25km grid is related to the 5km grid is simple – they are simply overlain in a GIS and change factors for the 25km grid cell are assigned to each individual 5km grid cell. To calculate the change factors for the catchment we then use the 25km grid cell nearest to the centroid of the selected area (we cannot average change factors from different 25km grid cells because they are not correlated) and then apply these to the point rainfall model fitted earlier to the catchment average rainfall statistics. We have added a sentence into the text to clarify this point "In the case of a small catchment, the CFs from the 25km grid cell nearest to the centroid of the selected area are used."

The second major remark is the use of the baseline runs. As I understand, the baseline runs were not run with any influence from the climate model, purely by the weather generator? This is good, but it does not tell you anything of the influence of the climate model over the same period. To create a control run, with which you compare against future scenario run, you would have to run simulations driven by the three scenarios over the period 1961-1990. As I understand, this information is not available from UKCP09 weather generator. This makes future scenarios very difficult to assess, since you are not comparing like with like. This is obvious from Figure 4 and table 1, where there is a larger difference between the baseline (which I would interpret as a proxy for observed values) and the scenarios (both in terms of mean and percentiles) than between the time periods of the scenarios themselves. Without any information about how the scenario-driven weather generator behaves over the control period, it is impossible to quantify the bias in the climate models. I understand that this is not possible to address fully within the study, but if it is not addressed, the results are not valid.

The referee is correct that the baseline runs were not run with any influence from the climate model. The weather generator is fitted purely using observed climate statistics from 1961-1990 and validated against observations. The way a future scenario is then generated is by applying change factors to these observed statistics. The change factors themselves are taken by calculating the difference (additive or multiplicative) between a climate model control (1961-1990) and future scenario (for whichever future time slice is the relevant case) (although in UKCP09 this is done in a more complicated way). This is the standard way that weather generators are run. We are therefore comparing like with like – you can think of a weather generator as a more sophisticated way of applying the 'perturbation' or 'delta change' downscaling method which normally scales observed time series by a 'delta change' (the ratio between the control and future run of a climate model for an observed quantity – normally the mean). In the case of the weather generator we are fitting this to the baseline "observed statistics" and then applying the change factors to the observed statistics (in this case we do not just change the mean but change 5 rainfall statistics and two temperature statistics) and then refitting it. The fit of the weather generator to the observed baseline was tested and was found to be good. Therefore the weather generator adequately simulates the baseline. We have changed the text slightly to try to clear up any confusion however "A set of 100 30-year simulations was also produced for the baseline period 1961-1990 against which to assess the future projections. The baseline period is fitted using observed statistics for 1961-1990. Note that variations in the baseline simulations represent solely the stochastic nature of the UKCP09 Weather Generator and can be thought of as an estimate of stochastic or 'natural climate variability' against which to compare the projected changes (which contain stochastic variability as well as changes consistent with the climate projections they are based upon). For each of the nine rainfall scenarios, 100 30-year daily simulations were then carried out with the CAESAR model, simulating daily sediment and water discharges"

The second point of the referee about the bias in the climate models themselves is a separate one. All climate models have biases in their control in relation to observations. Like many downscaling methods, a weather generator makes the implicit assumption that any bias in the control scenario of the climate model is the same for the future scenario – therefore although we would not believe the absolute values from the climate model we believe the projections of change. Hence, we use the climate model to give us 'change factors' between the control and future period for different climate statistics and apply these change factors to observed statistics for the baseline (which we believe as they are real!) This is a standard method in downscaling. For a review see Fowler et al. (2007) for example. However, this is valid point and we have now made this assumption clear by adding an extra sentence "This method makes the implicit assumption that any bias in the control scenario of the climate model is the same for the future scenario – therefore, although the absolute values from the climate model may be biased, we assume the projections of change are valid."

The third remark is on the treatment of extremes. Firstly, climate models were not constructed to give information on extreme events, rather on changes in the mean climate (where they are more robust), so any assessment of changes to extremes has to be taken with great care. The authors have addressed this in table 2 where the 5 and 95% percentiles are used as the lower and upper limits. However, in Fig 6, the high scenario indicates values of daily rainfall of almost 300 mm which results in maximum runoff almost 500 m3/s, which seems physically unrealistic, even under a changed climate. in figure 5, the different aspects of the high scenario are discussed, and the impact on the peaks around 700000 m3 are discussed, and this is a reasonable result, however, the impacts on the very high yields are very questionable given the extreme values of precipitation and runoff. *Figure 6 is a histogram and there is only one 300mm day rainfall event from 100, 30 year simulations. It has occurred once in 3000 years. Whilst rainfall events of this magnitude in the UK are rare, they have happened in Northern England – as evidenced by rainfall over Cumbria in 2009 which totalled 314mm in 24 hours* (Stewart, *et al.*, 2010)*. Given this rainfall, a flood discharge of 500m3/s is proportionate. Widespread geomorphic change was triggered by the floods from the Cumbria 2009 event including widespread aggradation, channel change and washing away of bridges. Furthermore, our interpretations in the rest of the paper are not based upon single days results – but groupings and percentiles as the reviewer identified above.*

The paper sets out to assess the a probabilistic framework of climate impact on sediment flow, but it lacks an uncertainty estimation from the different sources of the chain. It is a complex system, and the authors mention the sources of uncertainty, but there is no attempt to estimate the contribution of each source to the final uncertainty.

This is an excellent question and one that we are presently researching. It is simply beyond the scope of this paper to begin to address this question directly and there is presently a PhD student examining exactly this issue. We unfortunately have to limit the scope of the publication at some point.

Presentation The introduction is very well written and gives an excellent introduction to the subject and where this study fits in with the literature. However, there is no connection with the previous literature in the discussion. I would suggest a comparison with previous work to show the benefit and advances. The results from the sediment yield is presented and discussed before the rainfall and runoff, and I would suggest to do it the other way around. Sections 2, 3 and 4 are related to descriptions of methods and study area and could be put in one chapter with sub headings.

The novelty of this approach makes it difficult to compare to previous studies examining the impact of climate change on sediment yields. As shown in the introduction these can be grouped into soil erosion models, landscape evolution models and channel models. All of these show a general increase in sediment yield in response to climate changes and in the discussion we have made reference to this, by adding "Such increases in sediment associated with climatic increases in rainfall are broadly in line with soil erosion studies (Favis-Mortlock and Guerra, 1999; Chaplot, 2007) and channel models (Verhaar et al., 2010). " and "Previous applications of the CAESAR model (Coulthard and Macklin, 2001; Hancock and Coulthard 2011) have shown how sediment yield can be sensitive to increases in rainfall magnitude, but here we are able to determine what size events have the greatest impact.". We also face issues in comparison, as we are describing the impacts of climate change on an upland UK catchment – and other parts of the world may receive different – drier climates. Therefore we have termed the above sections in terms of the response to increased rainfall *magnitudes rather than climate changes. The main messages from the discussion are the 'geomorphic multiplier' – the increase in impact from rain > runoff > sediment – which we believe is a novel finding, one that is only quantifiable through a nested modelling approach as this. This is reflected in the second main point – the increase in very large sediment yield producing events.*

Figures Figure 2b is described as histogram, but it is a frequency-intensity diagram

This has been changed.

Figure 3 is very difficult to interpret because of all the experiments. The information is nicely presented in figure 4, so I would suggest deleting figure 3, or present a few cases as examples

We wanted a method to present all the data – allowing an appreciation of the variability that exists in the results across all the simulations. For this purpose we would like to keep Figure 3.

Figure 5. The black dots sometimes shadow the grey dots. I would suggest using different symbols to be able to show both .

These have been changed to clarify.

Figure 8 is not very clear to me, as well as the discussion, and this needs more explanation.

The text in this section has been edited and revised to make the explanation clearer.

References

Fowler, H.J., Blenkinsop, S. and Tebaldi, C. 2007. Linking climate change modelling to impact studies: recent advances in downscaling techniques for hydrological modelling. International Journal of Climatology, 27(12), 1547-1578.

Coulthard, T.J., Macklin, M.G., 2001. How sensitive are river systems to climate and land-use changes? A model-based evaluation. Journal of Quaternary Science 16, 347–351.

Macklin, M.G., Johnstone, E., Lewin, J., 2005. Pervasive and long-term forcing of Holocene river instability and flooding in Great Britain by centennial-scale climate change. The Holocene 15, 937–943.

Rumsby, B., 2000. Vertical accretion rates in fluvial systems: a comparison of volumetric and depth based estimates. Earth Surface Processes and Landforms 631, 617–631.

Stewart, L., Morris, D., Jones, D., Spencer, P., 2010. Extreme rainfall in Cumbria, November 2009-an assessment of storm rarity. Proceedints of the BHS Third International Symposium, Newcastle, 2010 1–5.