We thank Wouter Buytaert for his thorough reading of the manuscript and his constructive remarks on our paper. Below our responses to the general comments and suggestions for improvement on the specific comments (in bold)

General comments

This paper reports on a very ambitious attempt to predict

Response: we like to make clear that we do not predict but rather project impacts of drivers on water resources. In other words our focus is not on predicting future water resources but rather on understanding the relative magnitude of different drivers on water stress as mediated by the magnitude of the driver and its spatial pattern and distribution relative to water supply and demand centres

the impact of several physical and socio-economic drivers (land-use change, population growth, increase of industrial activities, climate change) on the water resources of the Peruvian Amazon using a global hydrological model (waterworld).

Waterworld is indeed very promising tool to analyse future scenarios of hydrological change in complex environments, in a more interactive process between scientists and stakeholders. There are several very interesting scientific questions here, which I think broadly fall in the following 3 categories:

1) User interaction with model scenarios and simulations;
2) Understanding the drivers of hydrological change in complex and data scarce environments;
3) Building models that make the best predictions in such environments.

Response: the focus of our paper is entirely on the second category: understanding the drivers of hydrological change in complex and data scarce environments. We have other papers that cover 1 and 3

However, the paper seems to hover indecisively between those three foci, and with that fails to make a real impact in any of them. I understand that the presented study is a model application, with the model itself being described in another publication by the senior author. As such, I do not expect this manuscript to focus on question (3). Nevertheless, the conclusions do make this claim (e.g., "combines the best globally available data with a broad and deep understanding of hydrological processes"). At best, the claim is unsubstantiated without a more through description of the model, and at worst a bit grandiloquent.

Response: Since we do not focus on question(3) and the model has been described in detail in Mulligan, 2013 we shall remove this sentence here.

Furthermore, claims that calibration is not necessary and that physical relations will hold in the future (p.570) are extremely controversial in hydrological modelling, and would also need careful justification. Several studies have for instance shown the large uncertainties and potential errors in remotely sensed data for that region (see e.g., Condom et al., 2011, Ward et al., 2011, Clarke et al., 2011). Similarly, representing hydrological processes is fraught with difficulties because of the dominance of poorly known hydrological processes such as groundwater (as also mentioned in the review of Christopher Scott, but see also for instance Zulkafli et al., 2012) and large-scale Amazonian wetlands that provide potentially very large and longterm storage (e.g., Guimberteau et al., 2012).

Response: Our focus is not on predicting water yields but rather on understanding the response to multiple drivers. We do not suggest that calibration is unnecessary. The model can be calibrated but since we are more focused on understanding the physical relationships between drivers and outcomes through hydrological processes, it would not be wise to calibrate the parameters of this physically based model, since that would negatively affect its physical basis. There are many examples throughout the literature of model experiments being of value even where the model is not validated and there are many cases where models cannot be validated (in scenario analysis or climate projection for example). WaterWorld has been in use since 2004 and has been applied by more than 1000 people at sites throughout the world. Over this time it has developed into a well tested and refined hydrological model. This paper is not about the model but rather the application of the model. Other papers discuss the model structure (Mulligan, 2012) and model testing (Mulligan and Burke, 2005; Bruijnzeel et al, 2011). See comment on groundwater in response to review of C. Scott.

In short, I fully agree that the use of satellite data and global model structures is a very sensible in data scarce regions, and can be useful as a first order estimation of hydrological behaviour. But this of course can entail big uncertainties and risks, which makes that the abovementioned claims need better support. For an introduction in some of the controversies related to the topic see the WRR commentary of Wood et al. (2011) and subsequent discussion initiated by Beven and Cloke (2011)). Again, this may well be beyond the scope of the project, but if so this needs to be reflected in the conclusions.

Response: agreed and we will add a section to the conclusions which will reflect on this.

Another really interesting aspect, highlighted by the manuscript, is the potential for user interaction provided by a web-based model implementation. However, this emphasized in the conclusions as a possibility, but not really substantiated in the manuscript itself. I concur with the review of Christopher Scott about the main problem being the lack of any user interaction or involvement presented in the manuscript.

Response: that is not the purpose of this paper. Other papers have and will explore these dimensions. The purpose of this paper is to explore multiple drivers, not user interactions. This will be reflected more clearly in the stated objectives.

But even without this, several other aspects could do with further deepening, especially because of the controversial nature of the topic. While some argue that hydrological models are too complex to be exposed directly to non-experts, this reviewer is not one of them (Buytaert et al., 2012). But that same publication highlights some of the many issues of interactive model simulation, with probably the main issue the need to communicate the impact of model assumptions, simplifications and uncertainties. Again, in a data scarce environment, the impact of the latter cannot be overstated (e.g., Beven et al., 2012).

Response: these issues are dealt with in a number of ways. a) We provide two model versions (1 and 2) that are different in complexity. Version 1 consists of the basic hydrological model while version 2 includes the snow and ice module, surface storage, water quality and wash erosion model. Furthermore, there are different interfaces for experts and non-experts with non-expert results presented in a much more simple way, denying access to the numbers, only colour scales given the uncertainty. b) the model can be run at different scales (1km and 1ha spatial resolution) c) the model allows for easy implementation with alternative inputs datasets and d) uncertainty in for example climate change scenarios can be assessed by running the model with individual GCMs as well as ensemble means and ensemble mean plus and minus the inter-model standard

deviation in the ensemble. This can be done within a few minutes for each simulation. We ask users to examine the impacts of data and model structure uncertainty using these tools and make it quick and easy for them to do so. It is a valuable lesson that all users get from running the system

One of the major risks of global model implementations is their false impression of data richness, which thus could be an incentive to decrease the pressure on local data collection.

Response: on the other hand they also make the best of the data that we do have, most of which goes unused because people do not have access or the knowhow to get into most model structures. Here they have access to it, can apply it, can see its limitations and can easily upload and use their own should they have access to a detailed monitoring programme or academic research project in their catchment.

This of course would be a perverse effect. On the contrary, using sensitivity and uncertainty analysis in an interactive simulation context can be a powerful tool to guide data collection investments to provide optimal benefit for local decision making. More in general, decision-makers often have high and unrealistic expectations regarding scientific knowledge and environmental models. The risk exists that they will see them as clear-cut cases for particular policies, or contrarily, they may lack confidence in models and results that they cannot reproduce themselves. All these and more questions come up when reading the manuscript. While of course not all can be solved, I agree with the other 2 reviewers that this needs much more attention. At least, the model being positioned as a policy tool, it should be possible to show how the model results are used to support policy decisions.

Response: that is not the focus of this paper. The paper is about exploring multiple drivers, not about policy makers using the model. We have other papers that cover those aspects.

Lastly, the tool holds great promise for understanding the potential variations in magnitude and the relative importance of the studied drivers, and as such to inform decisionmakers about priorities for policy making. This is what the authors focus most on in their description of methods and presentation of results. However, there is hardly any discussion of these results,

Response: agreed, this is what the paper is about so we will expand on discussion of the results

Again, I fully support the authors in their attempt to quantify the impact of the different drivers on hydrological change, and agree that such an attempt can be highly informative for policy (see for instance Buytaert and De Bievre, 2012, for a similar attempt to address climate change and population growth impacts). I even perceive an overemphasis on the impact of climate change at the cost of other drivers in current policies and results like these can help in correcting this. So what about for instance a synthesis figure (piecharts or barcharts) that presents the relative contribution of each driver to change for different parts of the basin (e.g., Andean highlands vs. Amazonian lowlands)? But then again, ideally, such figures should include some form of uncertainty estimation.

Response: agreed we shall add a figure with relative contribution of each driver to change for different parts of the basin.

The manuscript mentioned uncertainties several times, so it is really disappointing not to see more effort to quantify any of them. I again agree with the second reviewer that a sensitivity analysis would be a useful first step for this. A major disadvantage of sensitivity analysis is the often arbitrary range of variation.

Response: we will include a sensitivity analysis that will explore the relative change in water stress per unit change of each driver.

In several of the scenarios, however, the opportunity exists to use existing boundaries. For instance, while I agree that the GCM ensemble mean is a good choice as the best predictor (p 577, l17), it is also become standard practice to use the entire ensemble of GCMs (or at least a subset) to get to grips with the uncertainty of climate projections (e.g., Stainforth et al., 2007). This is especially the case for the tropical Andes, where climate projections show very large uncertainties, and there may be a high risk for systematic biases in the entire ensemble, for instance because of the inadequate representation of the topography (e.g., Buytaert et al., 2011). Given the availability of downscaled projections from CIAT, this should be quite straightforward.

Response: indeed, as mentioned in response above, the model includes all downscaled CIAT projections and simulations can be carried out with the mean and mean + and – the inter model standard deviation. This allows for analysis of the low and high end of projections as well as the mean. We propose to present these in the revised paper to more fully assess uncertainty

Similarly, the SRES scenarios (or the more recent RCPs) give ranges of population growth that could be used to constrain a sensitivity analysis.

To conclude, the manuscript holds great promise, but is currently very descriptive in the presentation of the tool and the scenarios. I think it could do much better in highlighting the scientific relevance. I hope the reflections above will encourage the authors to reflect on the direction of such elaboration, and I very much hope that it will be substantial.

Specific comments:

p568, l5. "particularly well suited to heterogeneous environments with little locally available data" -> how do you justify this?

Because it is grid based and focuses on spatially explicit variability unlike many semi-distributed approaches. We provide some data that is better than no data (which is what decision makers are very often faced with) but obviously not better than an instrumented catchment that could parameterise a more detailed model)

p568, l18. "See Mulligan et al. (2013) for a similar analysis for the entire Amazon Basin": So how does this manuscript complement the study by Mulligan et al. (2013)?

By focusing only on Peru and only on multiple drivers

p568, l26: Is 1ha really justified with the available data?

Yes, in mountainous environments where the terrain is highly variable and a 1km or worse resolution model would grossly oversimplify. It is the resolution of the data that is important here not the model as the same model is run at both resolutions. Our 1 hectare model uses Landsat derived tree, herb and bare cover and SRTM DEM, our 1km model uses MODIS veg cover and a 1km DEM.

I would be very reluctant to believe such results without some form of local validation. See for instance Buytaert et al., (2011) for an example of where high resolution models may give worse results than coarse resolution models. The discussion of Beven and Cloke (2012) is of course also very relevant here.

p570/ l12: The lack of possibility to calibrate a model highlights the need for some form of evaluation. I would be rather cautious in interpreting results of a model that is neither calibrated nor validated.
We can calibrate but we choose not to because we need to retain the physical realism of the model. Also it is not possible to validate these futures but as explained before the basic components of the model are widely used and well tested in a range of contexts as well as in Mulligan and Burke (2005); Bruijnzeel et al (2011). This is not to say that it should not be validated here but validating variables like water quality and water stress at these spatial scales is simply not feasible. That is not to say that exploration of the outcomes of multiple drivers is not useful, even if not validatable.

p574/16: how is irrigation accounted for?
It affects vegetation cover and is thus reflected in AET

p576: deforestation: even though the waterworld model is described in another publication, it would be very useful to have some description here of the hydrological parameterisation of the land-use classes. Otherwise it is impossible to interpret the potential impacts. Also, how is the land-use change model implemented? (l.2)?

Agreed, the model uses MODIS VCF (Hansen et al., 2006) fractional land cover types for herb, bare and tree cover. We will add a section on implementation of the land use model.

p578; population growth: as for the land-use change model, it would be very useful to have some details about how the population growth model is implemented, if only to assess the suitability of a (supposedly global) model for this particular region.

Agreed, we will provide more details on the implementation of this model

p578/20: of course water storage and transfers are already very common in the Andes...

p581:14: arguably, taking the mean is not a method to limit uncertainties, but only to identify a best estimator. The reference actually advocates the use of a model ensemble.
Will change sentence to reflect this

p582/5: it may be useful to state the main assumptions of the model beforehand (and maybe discuss the impact of those for the studied region as part of a section on model limitations).
Agreed, this will be added

p582/11: "hazard preparedness strategies" -> more typically referred to as adaptation strategies
Noted

p584/l8: Mulligan et al. (2013a): this reference seems wrong and should probably be that on p.586/13, which should be corrected to 2013c instead of 2013.
Will be corrected

p584/20: the surname of Waldo is "Lavado Casimiro", so should be corrected.
Will be corrected

p586/30: adn -> and

Will be corrected

References:

Bruijnzeel, L.A., Mulligan, M., Scatena, F.N., 2011. Hydrometeorology of tropical montane cloud forests: emerging patterns. Hydrological Processes, 25: 465-198.

Hansen, M. et al., 2006. Vegetation Continuous Fields MOD44B, 2001 percent tree cover, collection 3, University of Maryland, College Park, Maryland.

Mulligan, M., 2013. WaterWorld: a self-parameterising, physically-based model for application in data-poor but problem-rich environments globally. In press. Hydrology Research.

Mulligan, M., Burke, S., 2005. FIESTA Fog Interception for the Enhancement of Streamflow in Tropical Areas. Final technical report for AMBIOTEK contribution to DFID FRP R7991.