Editor comments and replies in **bold**

I'd like to thank the authors for uploading the revised version of their manuscript entitled „Global hydrological droughts in the 21st century under changing hydrological regime". The revisions are based on the reviewers' comments and the authors addressed these comments by providing a response letter.

With regards to the specific comment of Referee #1 on the selection of the GHM and its performance I would like to strengthen the importance of the first innovative aspect and discuss the model more prominently in the paper as recommended by Referee #1. The PCR-GLOBWB model is described in Section 2.1 and the advantages of using this model has been justified by its performance as described in earlier work (van Beek et al. 2011, Wada et al. 2012, 2013, 2014). There is no doubt about the model, however, there are two shortcomings related to the earlier work in combination with the application in this study, which I think require more detailed information:

First, the model has been validated and tested against 3600 GRDC observations but without any information on the number of stations where these observations were recorded. According to the references given in the text, the model has been evaluated in 23 major river basins (in case of drought deficit volume) and 26 major river basins (in case of monthly and annual mean flow). The authors should add this information to the text in order to further discuss this source of uncertainty.

First of all, we wish to thank the Editor for the thorough review on our revised manuscript. We really appreciate for your time that has been spent on the review. As suggested, we modified Section 2.1 (Model simulation of streamflow) accordingly and added further explanations of the model performance.

In brief, PCR-GLOBWB has been evaluated against 3613 GRDC stations with drainage areas larger than 2500 km^2, that is roughly equivalent to one grid cell (0.5 degree by 0.5 degree). These stations contain the long-term statistics of mean, minimum, and maximum discharge with sufficient data record (more than 10 years of monthly data), to evaluate our modelled streamflow. Because of the coarse spatial resolution of the model (0.5°), the upstream drainage area of some stations, particularly the smaller ones, cannot be represented accurately, thus they have not been included in our model evaluation. Notwithstanding, this data set provides a good starting point to evaluate the skill of the model to simulate discharge variations within and between years for varying catchment sizes and regions.

The validation of PCR-GLOBWB provides R^2 and the RMSE, biases compared to the observations from the GRDC stations in Van Beek et al. (2011) In addition, in Wada et al. (2014), the monthly streamflow has also been evaluated with R^2, slope, and Nash–Sutcliffe model efficiency coefficient.

As noted by the Editor the drought deficit volumes have only been evaluated at 23 major river basins (GRDC stations closest to the outlets) where drought has a major impact on the hydrology. We have stated this information in the revised manuscript.

Second, the performance measure R^2 is not a good one to evaluate the performance of a model to calculate deficit volumes based on daily time series as it is insensitive to systematic over- or underestimation. As a minimum, not only R^2 should be provided but also the slope of the regression line. In addition, Wada et al. 2013 have made a comparison of simulated deficit volumes per drought event against observed river discharge from selected GRDC stations in 23 river basins (R^2 = 0.75). However, this comparison is based on monthly rather than daily streamflow . Based on the previous work, for this study it would be of benefit if the authors showed the model performance based on daily streamflow for the Q90.

We have clarified the statistics used in the model validation in Van Beek et al. (2011) and Wada et al. (2014). Please also see our previous response. With regard to the validation of the daily model performance the Editor's comment is well taken. However, the authors believe that the current model performance based on the monthly streamflow validation provides sufficient information.

For this purpose, in Table 1 **a comparison is made for the Rhine River that shows that the model performance of PCR-GLOBWB does not deteriorate significantly when daily vales are used to simulate discharge. However, the authors believe that a full daily validation of the model goes beyond the scope of this paper.**

Although we used the simulated daily streamflow, the threshold is determined based on monthly aggregated daily values. The monthly threshold is determined and smoothed with a 30-day window. This is identical to, for example, the recent study of Prudhomme et al. (2014), which is considered as the best approach to determine drought characteristics for future projections from daily streamflow simulations. We compared the approach (used in this paper) to a purely daily approach where the threshold is determined only based on values from of that day (without smoothing using a 30-day window). For a 30 year period this would result in 30 values for the 30 year period. From those 30 values the daily threshold is determined, which is done for all 365 or 366 days of the year. This number is lower than the 900 observations that are available for the smoothened monthly approach. A high number of observations is especially important for extremes, so that is why the smoothened monthly approach is preferred. Furthermore Table 2 shows that the difference in drought characteristics is minimal and therefore we prefer the smoothened monthly approach for statistical more robust results.
Moreover, a comparison of the daily simulation with observations would be hampered by the fact that the hydrological simulations have been forced with GCM climate projections derived from RCP scenarios. Although the GCM climate forcings have been bias-corrected (Hempel et al., 2013), these coarse resolution simulations do not have enough skill to accurately reproduce historic hydrological conditions at a daily time resolution globally and hence a validation compared to daily observations (under the GCM climate forcings) would underestimate the performance of the model.

Another point we would like to make is that in this study we only look at relative impact on deficit volumes and do not look into the exact deficit volume. Even when biases occur in the model simulation (that cannot be retrieved from the R^2), these will be of no impact to results from this study (hereby assuming that bias are timeindependent).

Table 1 Performance of the model for the Rhine River at Lobith, comparison between daily and monthly.

Performance	Daily	Monthly

R^2	0.49	0.65
RMSE (m^3/s)	1300	710
Bias (m^3/s)	136	443

Table 2 Comparison between daily threshold approach and the smoothened monthly approached as used in our study.

Drought Characteristics	**Daily Threshold**	**Smoothened Monthly Threshold**
Number of events	118	119
Avg. Duration (d)	12.4	14.1
Avg. Deficit Volume (m^3/s)	2055	2157
Total Duration (d)	1458	1676
Total Deficit Volume (m^3/s)	242426	256636

As recommend by Referee #1, I invite the authors to further discuss their model results in the context of uncertainties related to the model structure (Sect. 4.3), i.e. the use of one GHM vs. an ensemble, as well as the uncertainties inherent to the model as provided in the author response.

We believe that PCR-GLOBWB model climatology is rather average compared to other GHMs. The work of Prudhomme et al. (2014) shows that the PCR-GLOBWB performance is rather comparable with other large scale hydrological models and hence a single model approach will not have a significant impact on the obtained results. We included this in the revised manuscript and made a reference to the Figure provide by Prudhomme et al. (2014), where PCR-GLOBWB is compared to other GHMs.

Regarding Section 4.2, the authors should avoid to compare the outcomes of their model simulations driven by RCP2.6 (RCP4.5) climate with the work of Wanders and van Lanen (2013) and Forzieri et al. (2014) who have used climate input based on SRES A2 and A1B, respectively.

We have adjusted the comparisons with Wanders and van Lanen (2013) and Forzieri et al. (2014) as suggested by the Editor. We removed any reference to the lower RCP 2.6 and 4.5.

Anonymous Referee #1

Received and published: 17 July 2014

General comments:

In this paper a transient variable threshold approach is proposed to adapt analysis of dry spells to longer-term changes in projected low flow regimes. The approach is aimed to analyse drought duration and deficit characteristic in future climate with respect to a changed regime, defined by monthly flow quantiles Q80 (80% exceedance probability). Current studies of the research group (e.g. Van Huijgenvoort et al. 2014) used the conventional (intransient) variable threshold approach where the results reflect change of drought characteristics with respect to current regime (using monthly Q80 values calculated from 30 years of observations. The paper at hand is aimed to provide a different view, by analysing anomalies with respect to a possible future low flow regime, derived from GCM projections.

While the paper is generally well written, it has a somewhat limited scope: It pursues a similar aim as previous studies (e.g. Van Huijgenvoort et al. 2014, Prudhomme et al. 2013), namely to assess the change of low flow and drought characteristics from global hydrological models (GHMs) forced with a set of GCMs and emission scenarios. While Van Huijgenvoort et al. 2014 used several GHMs, here only one GHM is used. This appears at a first sight as a weakness, unless the advantages of the applied GHM would be discussed more prominently in the paper, what I would certainly recommend.

There are two innovative aspects in this paper (which are in the current manuscript confounded and should be analysed and discussed more clearly). The first innovative aspect is that a rather novel PCR-GLOBWB model is used. Global models in general are afflicted with vast uncertainties and it would therefore interesting to compare the projections obtained here with other studies. However, this would require assessing the differences with respect to the same drought definition concept as in previous studies, i.e. the classical (intransient) threshold approach. This should yield into a discussion of similarity and differences, and the relative credibility of different models.

Thank you for your positive evaluation and thorough review. Your comments have really improved the quality of the manuscript. With regard to the use of PCR-GLOBWB model, we have added further explanations about the advantages of using this model in Section 2.1 (Model simulation of streamflow). In brief, the model has been applied for calculating drought events (e.g., Wada et al., 2013) and has been extensively validated for simulating drought deficit volumes and frequency compared to those derived from observed streamflow (at GRDC stations). Simulated minimum, average, and maximum streamflow has also been validated at the global scale and the model performance is good for simulating low flow conditions across regions with different climates. In Section 4 (Discussion) we already described an intercomparison with existing recent studies (Forzieri et al., 2014; Prudhomm et al., 2014), as well as uncertainties that are inherent to the model. The reviewer might have missed this in the context. The intercomparison shows that our outcome when using the classical (none-transient) threshold approach (*VTM*) agrees with the above-mentioned studies.

The second innovative aspect is the variable threshold approach. Here I am a bit in doubt about the interpretability of the dry spells characteristics with respect to a changed regime. While the classical approach is intended to assess future dry spells with respect to the current regime, the transient approach is intended to assess changes of anomalies with respect to a changed regime. So the future low flow regime may become more wet, but the anomalies could last longer ore have a larger volume under an increased threshold. But I am not sure for what application this may be useful (e.g. fish will be happy to be in a wetter environment (higher low flows!), but the statistics will tell them to be unlucky because they are at same time in a longer dry spell because the threshold was set higher for its evaluation. Not sure if this makes sense...). There should be a careful interpretation of the physical scope of this intransient threshold approach that makes clear what water resources management tasks can be served by this statistics. And if the new statistics are useful, a direct comparison on the results with the classical varying threshold method results is clearly indicated, so that one can learn about the differences of the two approaches.

The authors have included an extra section in the discussion of the paper describing the advantages and disadvantages of the transient threshold approach. For example, in what way fish is affected by the changes in hydrological regime. We make a distinction between water scarcity and drought and acknowledge that for some application the conventional threshold method (*VTM*) is more suited compared to our proposed transient approach(VTM_t). However, we argue that for most situations where the conventional threshold is better applicable we are looking more at water scarcity (imbalance between demand and supply) than drought. We hope the reviewer agrees with our vision on the applicability of the transient threshold as describe in Section 4.1.

To conclude, I think the paper has a greater potential as what it actually delivers in its present state. I would recommend that the scope of the paper is shaped according the two innovative aspects indicated above, to make it an interesting contribution to literature on projecting hydrological drought in future climate.

Minor comments:

P654, line 20: Give more specific information to what the R2 value belongs to (I assume total discharges, daily flows, and whole regime?). What is the performance for low flows then?

R2 value (0.9) belongs to respectively each of simulated mean, minimum, maximum, and seasonal flow evaluation (all at monthly time step), compared to those derived from GRDC observations. Each flow quantity was evaluated separately but the model performance was consistent with the different flow quantities. We have clarified this in the revised manuscript.

657, line 4: “AID(t) is the total area in drought at a given time t,”...is redundant; "at a given time (t)" may move to the sentence above.

We followed the suggestion by the reviewer.

Section 3.1: Seems to me that you are mixing low flows and threshold. Decide if you wish to address change in low flows Q90, or just thresholds for the drought analysis, and be then consistent throughout the paper.

We have adjusted Section 3.1 to only refer to low flows instead of a combination of low flows, and Q90 or thresholds. We would like to state that the low flow regime is derived from the threshold, so sometime the use of this term was still required.

References: Check reference style: Year of publication should appear at the end of each entry...

The references have been processed by ESD editorial office, apparently placing the page number where the reference occurs at the end of each entry. The reference style of the revised manuscript will be consistent with the reference style of ESD.

Anonymous Referee #2

Received and published: 11 September 2014

General comments

This paper is well written, well-structured and easy to follow. It addresses a relevant scientific question and falls within the scope of the journal. The objective of the paper is to globally address the impact of climate change on future hydrological drought by using a transient derived threshold for drought. This objective divides the paper into two parts: (1) a global hydrological model is used and future hydrological droughts are simulated by applying a combination of five GCMs and four emission scenarios (RCPs). The results show an increased drought duration and drought deficit volume in discharge. The uncertainty ranges are shown and the results are compared to other studies. Although this is not necessarily innovative in itself, the value lies in the rather high amount of up-to-date GCM-RCP combinations that have been used. Thereafter, (2) a transient method for defining the threshold for droughts is introduced and applied globally. By applying the new method where the threshold of drought changes over time, it is shown that simulated future droughts have a much shorter duration in comparison to when the non-transient conventional variable threshold (Van Loon and Van Lanen, 2012; Van Lanen et al., 2013) is applied. The authors discuss the differences in results, but the potential benefits of the proposed method remains somehow unclear. The authors suggest a dynamic definition of drought. Although this might be considered novel, it also raises an issue which is not discussed. With a definition of drought that changes over time, it will probably be more complicated to compare the magnitude of past, current and future drought events. To improve the paper I would therefore suggest discussing potential pitfalls when having a dynamic definition of drought. I would also like to see that the benefits of a dynamic drought definition are further elaborated upon.

Specific comments

- The manuscript shows how the drought duration and drought deficit volume is simulated to be greater for the non-transient approach in comparison to the dynamic, transient approach. This is valuable information since it shows how changes in the hydrological regime might influence the simulation of future drought duration. Nonetheless, I am not completely convinced by the transient approach and would suggest a clearer discussion in which way the transient approach is superior the non-transient. In my opinion the transient approach might also be slightly misleading. For example, decision makers might be interested in investing in hydropower. In order to understand how the runoff is about to change, they would probably want to compare the future situation in a direct relation to the current situation (non-transient analysis). In this example the transient approach would not allow a direct comparison between the time periods; instead it would underestimate the future drought duration. Hence, it would risk that the decision would be based on “false” information. With this in mind I would recommend elaborating on the potential weaknesses of the transient method.

The authors have included an extra section in the discussion of the paper describing the advantages and disadvantages of the transient threshold approach. We make a distinction

between water scarcity and drought and acknowledge that for some applications the conventional threshold is more suited compared to transient approach proposed in this study. However, we argue that for most situations where the conventional threshold is better applicable we are looking more at water scarcity (imbalance between demand and supply) than drought. We hope the reviewer agrees with our vision on the applicability of the transient threshold as describe in Section 4.1

- It is shown that different results are obtained with the transient and non-transient method. Still, I would like to see a clearer motivation why the transient method should be used. Furthermore, is the transient method suggested as "complete" or just applied in order to show that it is necessary to consider changes in the hydrological regime when addressing future droughts?

We do not claim the transient threshold approach to be superior over all other dought identification methods. However, the authors believe that we should adopt a more flexible approach when we identify drought conditions, i.e. the selected approach like the fixed threshold, the conventional variable threshold (*VTM*), transient variable threshold (VTM_t) or SPI, PDSI should depend on which impacted sector we study. Changes in future hydrological regime should definitely be included in the assessment of the potential impact of future hydrological drought, irrespective which drought identification is selected. The consideration of a gradually changing hydrological regime has not been quantified at a global scale before and we believe that our study provides an unique and comprehensive overview of the use of different threshold approaches. More details on the benefits and limitation of the proposed transient approach are given in Section 4.1

- Would it be advisable to apply the transient method also on other drought indices like the SPI, or when comparing current drought events with past ones?

We argue that it would also be good to apply the transient method to drought indices like the SPI or PDSI. For these drought indicators the climatology also changes (hence the probability distribution) due to climate changes. An effect that should be taken into account when considering changes in future climate. In this study we also have incorporated the trends in the threshold (e.g. Figure 2), which could also be applied to the SPI, where the changes in the precipitation distribution could be considered as well. We have added some thoughts on this issue in the Discussion Section 4.1.

- Page 658, line 14-18. How was it decided upon these thresholds for the robust decrease/increase etc.? Were they chosen arbitrary?

The chosen thresholds for robust decrease/increase are arbitrary. The motivation to select a minimum of 16 ensemble members was inspired by the fact that it would require that for all RCP on average 4 GCMs show the same directionality of change. When one GCM shows a different pattern one can assume that the majority could be seen as a robust and the one GCM as the outlier.

- Page 662, line 8-14. The VTMt results are presented before the VTM results. To facilitate the reading I would recommend presenting the methods the other way around throughout the paper.

That means; first the more traditional non-transient method, thereafter what changed when you applied the new one.

Since the focus of the paper is on the transient threshold approach and its novel aspects, the authors think that the order preferably should not be changed. Additionally, the differences between the thresholds have been highlighted in Figure 2. The impact of this Figure is first discussed in Section 3.1 and here clearly the difference between the two thresholds is described. Since the transient threshold is novel the authors prefer to first discuss the outcome of the new method and then compare with the more traditional method. This prevents that the reader will focus too much on the traditional non-transient approach. We hope the reviewer can agree with this vision on the structure of the manuscript.

- Page 664, line 3-4. "...which seems to be in line with their study." Here a high drought frequency is compared with high deficit volumes (extreme low-flows) and it is concluded that the result seem to agree. This seems to disagree with page 663, line 13-15 where a study by Wanders and Van Lanen (2013) shows lower drought frequency and increase deficit volume. Please clarify.

On Page 663 line 15-19 we state that there is some disagreement between the previous study of Wanders and Van Lanen (2013), i.e. in drought frequency, as a result of the use of different GCMs (Wanders and Van Lanen (2013): the CMIP3 simulation), and the current study (the CMIP5 simulation). Moreover, the results in the study of Wanders and Van Lanen (2013) are averaged over climate types and not for specific particular geographic regions. Forzieri et al. (2014) focus on Europe, while Wanders and Van Lanen (2013), focus on the global scale.

- Figure 1. Why is there such a fast drop in the transient threshold for the year 2075? I would expect it to be smoother if a running mean is used.

The fast drop in 2075 is caused by the fact that there is a series of drought events in a row which is not counteracted by extremely wet years. When a multi-year drought occurs the climatological average low flow regime could be significantly affected. On the contrary, a series of extreme wet years, would results in increase of the thresholds. Additionally, one can see that the threshold increases when the drought event of 2052 and 2061 are not anymore in the 30-year average. The normal low flow regime appears, i.e. the threshold increased again.

- Figure 4. The figure shows the area in drought for five major Koeppen-Geigner climatic regions. The manuscript suggests that the threshold for drought should be transient. To allow a more transparent comparison between the lines in figure 4 it would therefore be valuable to add a figure-line showing the changes in area of the Köppen-Geigner climate types under future climate. Alternatively, figure A1 could be extended. The changes could be derived based on a running mean of temperature and precipitation.

We agree with the reviewer that if we introduce a transient approach we should also dynamically adjust the location of the climate types across the globe and accordingly consider it in the Area in Drought (Figure 5, we trust that the reviewer means this figure with the area in drought). We have calculated the climate types for each location and made a simulation of the changes in climate types for the 21st century. An animation of the changes is included in the supplementary material. Also Figure 4 has been adjusted to give the AID that considers the changing climate types at a

location. Furthermore Figure A2 has been added to show the major climate types in 2099 based on all RCP scenarios. Furthermore, two extra sections have been added to explain the methodology (Sections 2.5 and 3.4).

Technical corrections/suggestions

Page 651, line 6: “The 2011 drought in the Horn of Africa caused large famine across the region...”

We followed the reviewer’s suggestion.

Page 651, line 8: “Drought, heat waves and forest fires caused almost 80.000 deaths in Europe”, under which period, 1998-2009?

These 80.00 deaths are only for the year 2003.

Page 651, line 11: some -> certain

We followed the reviewer’s suggestion.

Page 651, line 18: “precipitation and/or temperature, which also propagate to reduced soil moisture”

We followed the reviewer’s suggestion.

Page 652, line 1: “Furthermore, Forzieri et al. (2014a) only assesses future drought for one continent (i.e. Europe)”. It should be mentioned that the resolution is higher (ca.25km).

We followed the reviewer’s suggestion.

Page 653, line 1: erase “used in this study”

We followed the reviewer’s suggestion.

Page 658, line 7: “130 years of observed and simulated discharge”

We modified the manuscript to “130 years of simulated river discharge”, since we don’t have 130 year records of observed future discharge.

Page 658, line 13: “significant ($p < 0.05$) trends were taken...”

We followed the reviewer’s suggestion.

Page 661, line 7-8: do you mean Fig 3-4?

We use Figure 2 to show the trends in low flow regime and Figure 3 to show the trends in transient threshold drought characteristic. By accident Figure 3 and 4 were swapped, now it should be correct.

Page 661, line 10: “Precipitation totals for these regions show an increase of 30–100mm-1 year for the period 1971–2000 compared to 2070–2099...” Unclear, isn’t that a decrease in precipitation?

The periods have been interchanges. We have corrected the manuscript to "Precipitation totals for these regions show an increase of 30–100mm-1 year for the period 2070–2099 compared to 1971–2000...".

Page 661, line 10 and 24: use "annual precipitation" rather than mm year-1

We followed the reviewer's suggestion.

Page 662, line 2: occurs

We followed the reviewer's suggestion.

Page 662, line 17: remove "slightly"

We followed the reviewer's suggestion.

References:

Forzieri, G., Feyen, L., Rojas, R., Flörke, M., Wimmer, F., and Bianchi, A.: Ensemble projections of future streamflow droughts in Europe, Hydrol. Earth Syst. Sci., 18, 85–108, doi:10.5194/hess-18-85-2014, 2014.

Prudhomme, C., Giuntoli, I., Robinson, E. L., Clark, D. B., Arnell, N. W., Dankers, R., Fekete, B. M., Franssen, W., Gerten, D., Gosling, S. N., Hagemann, S., Hannah, D. M., Kim, H., Masaki, Y., Satoh, Y., Stacke, T.,Wada, Y., and Wisser, D.: Hydrological droughts in the 21st century, hotspots and uncertainties from a global multimodel ensemble experiment, Proc. Natl. Acad. Sci., 111, 3262 – 3267, doi:10.1073/pnas.1222473110, 2014.

Wada, Y., van Beek, L. P. H., Wanders, N., and Bierkens, M. F. P.: Human water consumption intensifies hydrological drought worldwide, Environ. Res. Lett., 8, 034036, doi:10.1088/1748-9326/8/3/034036, 2013.