We are thankful to Referee # 2 for the critical review and very useful comments. We are glad to read that the Referee considers our basin-wise investigations as a valuable approach and similarly the study region as a worthy target area for such kind of analysis. The main issues raised by Referee # 2 are: 1) to exclude the P+E diagnostic indicator because it is redundant and not very useful; 2) the little relevance of comparing the Indus River discharge into sea with the models' simulated integrated runoff; 3) to attain a more comprehensive and synthetic description of the models' outputs, presenting e.g. results in terms of ensemble quantities etc. We mostly agree with the major remarks provided by the referee.

Our responses to the Referee's specific comments are given below:

Specific Response

1. **Unfortunately it does not become clear what is really new in this study.**

   We thank the Referee # 2 for providing us an opportunity to clearly state what is novel in our study. First of all, our study first time provides the investigations on whether the water balance of the CMIP3 climate models is closed for the major river basins of South and South East Asia, pointing out inconsistencies of the individual models in detail, if any. This issue is of relevance for understanding to what extent the representation of various hydro-climatic processes is adequate over the region. The study states the different behaviors of the models regarding their representation of the hydrological cycle over one of the critical region of the World in terms of effectively simulating the hydro-climatological characteristics by climate models.

   Secondly, though the study mentioned by the Referee # 2 i.e. Milly et al., (2005), and Nohara et al., (2006) and others investigate/project the runoff quantities, more often, based on the ensemble quantities, these studies do not discuss individual models' performances. Thus, keeping in view the non-uniform response among models regarding the projected sign of change, e.g. for South East Asian Rivers (Arora and Boer, 2001; Arnell, 1999), and the need for the specific model's performance by the communities involved in climate downscaling and various impact assessment applications at a local scale, these kind of studies can't be of much relevance. In this respect, presenting individual models' behaviors to fulfill the needs of the targeted end user communities is an important aspect of our study.

   Furthermore, the studies providing observed/simulated water balance and runoff quantities are usually either global – do not provide catchment scale characteristics of the hydrological cycle - or always missing one or the other South and South East Asian river basins (Arnell, 1999; Nijssen et al., 2001; Douville et al., 2002; Manabe et al., 2004; Hagemann et al., 2005; Nohara et al., 2006; Pan et al., 2011). Thus, the present study fits in quite well, filling in this analysis gap by adequately addressing the basin scale hydrological cycle representation by the individual/cluster of models, for all the major river basins of South and South East Asia region. The novelty of the study will be more clearly presented in the new version of the paper.

2. **I wonder why CMIP3 models are investigated, and not CMIP5 models.**

We have chosen the CMIP3 data set on purpose to start our investigations in a quite systematic fashion which will be extended to CMIP5 data set, providing basin-wise performance details of the models, and that how these models foresee the changes in the hydrological cycle under warmer climates. Keeping in mind that our investigation deals with quasi-regional scale properties of the hydrological cycle, our present focus on CMIP3 climate models is rather relevant when we consider local downscaling communities (like CLASH project under which this study is conducted) and operational organizations using regional climate models which are, at the moment, nested into CMIP3 GCMs, for instance, Pakistan Meteorological Department (PMD) and Global Change Impact Studies Centre (GCISC), Pakistan, Department of Hydrology and Meteorology (DHM), Nepal, Indian Institute for Tropical Meteorology (IIT), India, National Climate Center of Thailand (NCCT), Thailand and so on. It is tempting but sometimes incorrect to assume that the latest version of a model is the most relevant for everyone and for all purposes.

Obviously, we are aware of the CMIP5 data availability and the extensive development undergone for the new generation of climate models, introducing higher resolutions, atmosphere and land use and vegetation interaction, detailed aerosols treatment, carbon cycle, etc. (Taylor et al., 2011). Apart from these substantial developments, CMIP5 models are reported to feature similar projected changes in the global temperature, consistent spatial patterns of temperature and precipitation change and no significant reduction in models' spread as compared to CMIP3 climate models (Knutti and Sedlá˘cek, 2012). Nonetheless, it is worth interesting to investigate how the new generation climate models represent the hydrological cycle over our studied region at a river basin scale. Therefore, as an extended part of our systematic investigations, clearly mentioned in the paper (p. 148 – line 12-15), we have already started analyzing the output of CMIP5 climate models. The results will be reported elsewhere, and soon. The CMIP3 climate model analysis will then serve as a benchmark for our future work, providing an opportunity to see how the newly introduced features and enhanced processes now implemented in several CMIP5 climate models have impacted the representation of the hydrological cycle over the region. We will more clearly state the above cited reasons in the Introduction and Data sections of the revised paper.

3. **CMIP3 models are designated as 'present day climate models', which is actually not correct anymore as this is now true for the CMIP5 models.**

We agree with the Referee # 2 and we will replace the 'present day climate models' with 'CMIP3 climate models' in the revised version of the manuscript.

4. **Many results on the global hydrological cycle and its projected changes simulated by the CMIP3 models are already published and, e.g., summarized in AR4 of IPCC (Solomon et al. 2007). For example, Dai (2006) considered various precipitation characteristics in 18 CMIP3 models using global maps, Milly et al. (2005) considered future changes in runoff projected by CMIP3 models. The same also applies to studies of South Asian monsoon precipitation simulated by the CMIP3 models (e.g. Annamalai et al. 2007). Well, in this respect, the thorough analysis and discussion of reasons for common CMIP3 model biases over the South Asian region would still provide some new insights, but this is not done in the present study.**

As stated earlier, the mentioned studies (Milly et al., 2005; Annamalai et al., 2007) and other present results/projections about the specific aspects of the hydrological cycle, which are not necessarily the most relevant for the regional climate modelers, impact assessment scientific communities and water managers because of either their global/regional (not catchment) scale or widely adopted ensemble approach. As far as investigations of CMIP3 model biases over the South Asian region are concerned, many studies (Kripalani et al., 1997; Kang et al., 2002; Annamalai et al., 2007; Lin et al., 2008) are performed in this regard, associating the model biases with the inaccurate representation of the monsoon and its interaction with the westerly disturbances, which is clearly mentioned in the paper (P112, line 17-18). We have cited the findings by Boos and Hurley (2012) linking the misrepresentation of Hindu kush Karakoram Himalaya (HKH) topography by CMIP3 and CMIP5 models with biases in the dynamics and thermodynamics of the monsoonal circulation (P112 – line 20-24), which in our point of view, is the main cause of the negative monsoon anomalies and its delayed onset over the study region. In addition to that the analysis presented in this paper regarding the water balance inconsistencies of the models is also one of the possible causes of biases in their simulations of regional circulations. We will try to expand this discussion in the revised paper.

5. **On p. 121 line 1-2, it is proposed 'to understand the range of climate projections in the later part of the XXI and XXII centuries', but I couldn't find explanations with regard to this objective.**

   We will replace the word 'understand' with the word 'evaluate'.

6. **Sometimes the authors state a kind of ranking of model behavior between the basins (e.g. p. 133 – line 14-17). What's the scientific value of such a ranking?**

   We do not propose to rank models; we just wish to present the individual models' outputs. This, we believe, is indeed relevant. In any case, this comment becomes irrelevant in the light of our agreement to exclude the whole P+E part of our analysis in the revised submission (kindly see our response to the comment # 12).

7. **The reasons of the different model behaviours are of interest, but no explanation is given here.**

   It is well known that the different model behaviors are characterized by their structural differences, implemented physical based processes and parameterization schemes. Model specific characteristics are generally discussed in their respective documentation/ publications. Investigating different model behaviors for the considered hydrological quantities over all four study basins requires further analysis of dynamic and thermodynamic processes, land-surface parameterization schemes and relevant quantities. This is beyond the scope of the present study and secondly, such an analysis could span over bunch of studies. Hence, sticking to the definite aims of our study, we disagree with referee regarding incorporating this analysis in the paper. However, it is relevant to mention here that in our next study we have decomposed our analysis on an intra-annual time scale, where we will be presenting behavior of individual models in terms of their skill in reproducing the general characteristics of the large scale circulations performing over the region.

8. **Differences in the precipitation characteristics are explained by different climatic conditions (p142 line 6++), but this is known fact and not related to the models. Certainly the models should simulate the different regimes. Biases in hydrological variables are explained by biases in precipitation, but a general explanation for the latter is not given, nor for the large inter-model ranges, with one exception.**

   General description for biases in precipitation is discussed in the same section (P142 – line 20-22) and (P143 – line 1-6) and now (P143 – line 21+).

9. **On P143 – line 16++. Here, some monsoon features are explained, but unfortunately I don't get the connection with the intermodel spread and the model biases from the current text.**

   Here we will add the following text at (P143 – line 21)

   "Unfortunately, the phenomenon is not realistically represented in CMIP3 climate models due to their misrepresentation of the real topography, resulting in the negative monsoonal precipitation anomalies (Boos and Hurley, 2012). Further, as we know that an accurate representation of real topography by the models is mainly constrained by their structural characteristics (e.g. resolution), these structural differences among models are possibly one of the major causes of the inter-model spread of precipitation regime over the region."

10. **The present study is merely descriptive, and it very explicitly describes the simulation characteristics for each individual CMIP3 model for each river basin. In addition, it provides a lot figures that look all very similar. The authors need to find a way to summarize and present their results in more condensed form. I suggest focusing on important results and summarizing the important results for each basin in a concise way. The authors should avoid the description of every single model behaviour (science is not a beauty contest). Except for considering specific outliers of interest, they should use summarizing, more general descriptions such as the ensemble mean, ensemble spread and some measure of clustering. With respect to the latter, the authors provide a very interesting result that gets almost lost in the current version of the paper (see Sect. 5.1). They note that the ensemble means over the Brahmaputra and Mekong basins do not resemble any single model result. This may mean that the ensemble mean does not represent a realistic solution of climate over this region. In this respect, defining a measure that may indicate this failure of an ensemble mean would be a large step forward. In this respect, it is valuable to note, that a n umber of models clusters for specific basins and variables, but usually not which models exactly agree with each other (such as done, e.g., p. 132 – line 13-15).**

    We wish to clarify that the present study differentiates itself from most of the existing literature by deliberately avoiding the presentations of results in terms of ensemble quantities which are given in the paper for indicative purpose only. In fact, there is no rigorous way to condense our findings as averaging different models outputs is not a well-defined operation in any rigorous statistical sense. We believe that there is actually no strong reason to believe that the weighted multi-model ensemble estimates necessarily outperform any single best model (although this has become a standard practice within the climate's science community) - mainly due to the huge structural and physical differences among the ensemble members and that these members do not belong to the same probability space - nor there exists any well-defined mechanism to weight the individual models according to their performances at all. Many scientists have discussed this matter. Under such circumstances, we are also not convinced to summarize our findings in terms of ensemble quantities, which can indeed be misleading and be not truly representative of the respective datasets. We have extensively discussed this in previous publications (Lucarini et al. 2007 Climate Dynamics, Lucarini et al. 2008, JGR; this latter one focused on the hydrology of the Danube).

We maintain that providing individual models' outputs is of relevance. Science is indeed not a beauty contest, but we think that providing users with some measure of the performance of individual model is more informative than providing information about the ensemble mean, which is just a statistical construction, which does not correspond to any actual model and has no rigorous justification in Statistics. Moreover, in downscaling applications, RCMs are nested into specific GCMs, and it is crucial for user to have an idea of whether the model they are using has biases at reasonably large scales.

We can surely define, following the suggestion by the Referee # 2, a quantity indicating roughly the degree of agreement of the model's outputs. A good candidate could be the ratio of the ensemble spread and of the ensemble mean.

11. **The use of some hydrological variables is not common or does not seem to provide new insights. The authors use the term water balance (starting already in the abstract) to quantify P-E, while the correct hydrological usage of this term is P-E-R = dW/Dt. For long-term climatological means (dW/dt ! 0), this becomes P-E-R = 0. If this equation is valid for an individual model its water balance is closed. (This is also investigated in Sect. 4.1.3, but using the misleading terminology.)**

    We fully agree with the Referee # 2 that using the expression water balance when indicating 'P-E' can be misleading. Therefore, we will replace the 'water balance' with 'P-E' wherever it is used in terms of P-E in the next version of our paper.

12. **In addition the strength of the hydrological cycle, P+E, is considered. But all biases or changes in this quantity are related to biases or changes in precipitation and evaporation that are also considered. In this way, considering P and E separately is sufficient, and considering the constructed quantity of P+E is obsolete and does not gain any valuable additional information.**

    We agree with the Referee # 2 that it is not worthwhile to discuss also P+E. It will be removed in the revised version.

13. **Some of the analyses for the current climate seem to ignore knowledge about the basins itself. In this respect it is very inappropriate to compare the simulated runoff over the Indus basin to observed Indus river discharge. The latter is strongly anthropogenically influenced due to the large irrigation activities over the Indus basin (Karim and Veizer 2002). This means that large amounts of water are removed from the Indus River and they are evaporated over the irrigated areas. Consequently, as none of the CMIP3 GCMs includes irrigation, it cannot be expected that any of the GCMs is simulating the observed discharge values. Thus, it is not surprising that almost all GCMs overestimate P-E = R over this region.**

    Though we have stated that agriculture is the major consumer of the supplied water (P116, line 2), we thank the Referee # 2 for drawing our attention to discuss the main reason of Indus basin P-E and R overestimation by most of the models. To address this comment, we have decided to compare the simulated runoff of the Indus basin with the "natural" discharge, which is obtained by adding the total diverted volume within Indus basin to the volume discharged into the sea. This will be explicitly stated in the caption of the figures (2a & 8a) and in the text. The comparison between models and data will be refocused on this quantity.

    According to the APHRODITE precipitation dataset (Yatagai et al., 2012) climatology (1961-2000), Indus Basin precipitation accounts for 364 $mmyr^{-1}$ while its mean observed discharge into sea is around 40 $mmyr^{-1}$ runoff equivalent (Karim and Veizer, 2002). The total water diversion (170 $mmyr^{-1}$) from the Indus basin is around 80% of its mean surface water availability (210 $mmyr^{-1}$) as reported by Laghari et al., (2012). Hence, assuming no diversion within Indus basin, the net volume of the natural discharge into sea is roughly estimated to be around 210 $mmyr^{-1}$ runoff equivalent (170$mmyr^{-1}$ total diverted runoff within basin + 40$mmyr^{-1}$ runoff equivalent of the observed discharge into sea). The estimated natural discharge will additionally be compared with the simulated runoff and will be discussed in the text accordingly. We will also explicitly describe the fact that CMIP3 climate models do not feature irrigation.

14. **For the Mekong an additional problem occurs due to the relatively narrow structure in the Northern part of the basin. Here, the coarse scale resolution GCMs certainly have problems to adequately simulate the fine-scale precipitation over this area. This reason for biases is neglected in the present study.**

    We thank the Referee # 2 for directing us towards one of the possible causes of coarse resolution climate model biases for Mekong basin. The APHRODITE Precipitation dataset (Yatagai et al., 2012) reveals that the narrow gorge of northern part of Mekong basin receives the smallest amount of precipitation if compared to the remaining part of the basin. The major portion of precipitation does indeed fall over the southern part of the basin (below 23^{o}N). Such pattern has been observed also in the GCMs considered in our study. Moreover, the northern part of the basin features a very small surface area, hence, the calculated basin wide integrated quantities are not much affected by the potential biases in this region. Nonetheless, we will briefly mention this fact as one of the possible reason of the biases in the revised version.

15. The English needs major improvements. For future versions of the manuscript, I suggest proof reading by a native speaker.

We will get our manuscript proofread by a native speaker before its resubmission.