Response to Reviewer's Comments : Anonymous Referee #2

GENERAL COMMENTS

(Referee Comment, RC) The author presents a study aimed at improving model performance in a regionalization/ungauged basin context through a calibration approach that considers wet/dry catchment states. Unfortunately, the material presented is very similar to previously published material by the author and others (lacks novelty) **(Please see AC I below)**. The outlined study attempts to draw conclusions based on a small sample size (<10 catchments) and mixed results **(Please see AC II below)**. Additionally, the idea of fixing stated model structural errors/deficiencies by manipulating the calibration process is not particularly sound **(Please see AC III below)**. Such an approach will always be susceptible to failure under alternate conditions (i.e., the validation period) – something which was not considered at all in this study **(Please see AC IV below)**.

(Author Comment for General Comments, AC I)

This study investigates the possibility to improve the hydrological model parameter regionalization approach through rainfall-runoff model calibration for the different parts of flow regime, particularly focusing on model behavior during dry periods.

The current paper is part of the same line of works by Kim and Lee (2014a, b); such as rainfall-runoff model, study sites and multi-objective optimization algorithm, but it has sufficient difference from such works in terms of research objectives. The work presented by Kim and Lee (2014b) addresses the use of regional regression of model parameters (through multi-objective calibration), but the seasonal calibration is not adopted; whereas Kim and Lee (2014a) proposes the use of a seasonal calibration (similar to the focus on the dry season applied to the present study), but does not carry out a regional regression of the model parameters (but only applies calibration of single catchments).

Kim and Lee (2014b) assessed the efficiency of the regionalisation approach based on a multiple objective calibration technique to rainfall–runoff modelling under the catchment conditions of the Republic of Korea. They concluded that the regional model with the multiple objective approach led to improved hydrological simulations in ungauged catchments over a single-objective approach, but there was still a flow-dependent bias (i.e. a tendency towards the behaviour of underestimating and overestimating high and low flows for the wet and dry periods, respectively) in the runoff simulations. The present paper builds on their analysis and examines the relative performance of regionalisation methods. The work presented in this paper is a methodological advancement of the work by Kim and Lee (2014b) in regionalisation studies and goes beyond Kim and Lee (2014b) in three important aspects.

First, the variability (or consistency) of hydrologic response in model performance and parameter values and the NIREs (non-parametric impulse response estimates) have been implemented to investigate problems associated with the model accuracy under the different flow regime, which might be accumulated through regionalisation processes (i.e., leading to influences on the identification of relationships between the calibrated parameter values and the catchment characteristics as well as runoff predictions in ungauged catchments). Second, the calibration strategy (splitting the historical time series into wet and dry periods and calibrating data segments only for the dry period separately on the LRM (linear routing module) parameters because of a sufficiently high degree of similarity between the LRM parameter values for the wet period and those for the whole period) was established to improve the identification of relationships, according to an identification of deficiencies on model structure suitability for the different flow regime. The idea on seasonal calibration (Kim and Lee, 2014a) was accepted here. Third, Kim and Lee (2014b) concluded that a flow-dependent bias (i.e. underestimation and overestimation for the wet and dry periods, respectively) in the regional model might be reduced through modifying the model structure. However, the present study demonstrated the flow-dependent bias for the dry period in the regional model can be improved with an advanced calibration approach through improving the identification of the relationships between the catchment characteristics and the calibrated model parameters 'without modification of the model structure'.

Such issues were already discussed with the editors (handling editor: Dr. Elena Toth; handling executive editor: Prof. Dr. Erwin Zehe) through the editor decision (26 May 2015). The editors concluded that the present manuscript has sufficient difference from the previous works to allow being considered for publication and required to add an addition explanation on that in the manuscript. There was a minor revision on the introduction to make explicit what is in common with the two previous studies and what is novel in the submission after the discussion with the handling editors. Furthermore, more details on these issues (discussed above) will be added in the introduction and the conclusion sections in a next draft in order to clarify the advancements from the previous works.

(Author Comment for General Comments, AC II)

Please see the Author Comment 5 (AC 5) below for details on an issue related to a sample size.

(Author Comment for General Comments, AC III)

Please see the Author Comment 2 (AC 2) for an issue regarding a choice of modelling approach (i.e., improvement of calibration method and modification of model structure).

(Author Comment for General Comments, AC IV)

A model assessment scheme commonly used was proposed by Klemes (1986), involving a four-level test scheme (calibration and validation) involving simple split-sampling (stationary condition at the same catchment), differential split-sampling (non-stationary condition at the same catchment), simple proxy-catchment (stationary condition at different catchments), and differential proxy-catchment tests (non-stationary condition at different catchments). The term stationary was used to denote physical conditions that do not appreciably change with time. The simple split-sampling test uses subsets of the time-series record for calibration and validation. On the other hand, the differential split-sampling test involves calibration and validation under different meteorological conditions (e.g. dry and wet periods). The simple proxy-catchment test can be used to show the general validity for different catchments. The differential proxy-catchment test is a combined approach, incorporating differential split-sampling and simple proxy-catchment tests. The split-sampling test is appropriate when the available record is sufficiently long to split into two equal parts. However, the other level tests should be considered if the available record cannot be meaningfully split (Klemes, 1986). The simple proxy-catchment test should be required when an available streamflow record in a catchment is not adequate for the simple split-sampling test and also be required as a basic test for the general model transposability such as ‘regionalization studies’ (Klemes, 1986).

The level test similar to the simple proxy-catchment test (Wagener and Wheater, 2006) was implemented in this regionalization study (Wagener and Wheater (2006) validated their regional models using this approach). The calibrated parameter values for the traditional and the refined approaches (in the nine calibration catchments) were evaluated in the validation catchments (i.e., the two catchments that were excluded in deriving the regional models) through the comparative assessment of the accuracy of predictions of the regional models derived from the two approaches: the traditional and the refined approaches. This will be discussed more in the next draft.

In addition, an appropriate length of record needed in model calibration was considered in order to obtain time-invariant parameter values and statistics with a minimum amount of data and reasonably high model performance. At least 8 years or more length of record was needed to obtain a reasonable catchment response in calibration (i.e., to stabilize performance statistics and detect parameter consistency in calibration, which is the important issue in regionalization studies) in the IHACRES rainfall-runoff model (Kim et al., 2008). Kim et al. (2008) investigated the convergence of the model parameters to time-invariant values with increasing length of calibration period using the mean, standard deviations (SDs) and coefficients of variation of the various model performance indicators and parameter values. They proposed that an 8-year calibration period is appropriate for obtaining a reasonable catchment response, yielding stable and reasonably high model performance and reduced variation in parameter values over time for the IHACRES. In this study, an 8-year calibration period from 2003 to 2010 was used based on the previous study. However, available streamflow records are not sufficiently long enough to split into two equal parts at some catchments such as Hoengseong, Nam River, Miryang, Yongdam and Buan catchments (Table1 in the manuscript). This is another reason for the use of the level test similar to the simple proxy-catchment test in this study. Details on an evaluation approach for the calibrated parameter values caused by data limitation will also be added more in the discussion part (6. Summary and conclusion) in the next draft.

SPECIFIC COMMENTS

1. **(RC)** P7059, L24-26: It is necessary to discuss this in depth. Particularly, there have been several recent studies which utilize hierarchical Bayesian approaches to address regional calibration. See Bulygina et al., 2012 (doi:10.1029/2011WR011207) and Smith et al., 2013 (doi:10.1002/2013WR015079).

(Author Comment for Specific Comments, AC 1)

A next draft will include recent regionalization studies utilizing a Bayesian approach: research on integrating three different types of information into physics-based model parameter estimation (Bulygina et al., 2012) and introducing a new technique based on the hierarchical Bayes empirical Bayes methodology (Smith et al., 2013) in ungauged catchments. The next draft will also include additional literature review of rainfall-runoff model parameter regionalization research over recent years: a model independent approach to streamflow prediction in ungauged basins based on empirical evidence of relationships between watershed structure, climate and watershed response behavior (Yadav et al., 2007); an assessment of the effectiveness of hydrological model parameter regionalization approaches (Oudin et al., 2008;Merz and Blöschl, 2004;McIntyre et al., 2005;Parajka et al., 2005;Kay et al., 2006;Young, 2006;Wale et al., 2009;Zhang and Chiew, 2009); a multiscale parameter regionalization technique (Samaniego et al., 2010;Kumar et al., 2013). Besides, literature review regarding studies on a sequential regionalisation (Lamb et al., 2000;Wagener and Wheater, 2006), a regional calibration approach in regionalization (Hundecha and Bardossy, 2004;Hundecha et al., 2008), a regionalization approach combining an iterative regional calibration (Parajka et al., 2007) will be presented in the next draft.

2. **(RC)** P7061, L8-11: If the model is known to be deficient, why not correct the model deficiency itself rather than using calibration restrictions to force a different model behavior?

(Author Comment for Specific Comments, AC 2)

There are many sources of uncertainty that affect the regionalization process of the regression approach, including uncertainties due to parameter estimation, model structure and data used in the rainfall-runoff model and the regional model. The uncertainty in the calibration parameter values influences the identification of the relationship between the model parameters and the catchment attributes (Beven and Freer, 2001;Merz and Blöschl, 2004). The uncertainty in parameter estimation is significantly increased when the 'best' parameter set cannot be uniquely located through model calibration (i.e. the equifinality problem of Beven and Freer (2001)).

Such problems with parameter identification can be minimized by limiting the number of parameters to be estimated and by using information about the system to limit the range of possible values (Dunn, 1999;Seibert and McDonnell, 2002;Koivusalo and Kokkonen, 2003;Uhlenbrook and Sieber, 2005). Parsimonious model structures have been proposed to overcome the limitation of the high number interactions between parameters in regionalization studies by reducing model complexity (Wagener and Wheater, 2006;Hundecha et al., 2008;Sefton and Howarth, 1998;Post and Jakeman, 1999;Kokkonen et al., 2003). In this regionalization study, the parsimonious model structure (the IHACRES model), requiring minimal input data (see the third paragraph in AC 6 for the data limitation in Korea), was adopted to avoid adverse effects arising from model complexity, over-parameterization and data requirements. The IHACRES has been applied to a variety of catchments of different sizes and under different climate conditions (Jakeman and Littlewood, 1993;Schreider et al., 1995;Post et al., 1996;Schreider et al., 1997;Ye et al., 1997;Letcher et al., 2001;Littlewood, 2002;Croke et al., 2004). It was also widely used in regionalization studies (Sefton and Howarth, 1998;Post and Jakeman, 1999;Kokkonen et al., 2003).

As the reviewer said, the model deficiency can be improved through modification of the model structure leading to an increase in model complexity. However, there are some issues related to complexity of models that reduce their effectiveness. An increase in model complexity not only means an inevitable increase in data requirements, but also easily results in ill-conditioning and non-identifiable parameters (Dietrich et al., 1993). In particular, the increased complexity in the model structure may lead to greater variability if the parameter interaction is too great. This variability in estimated parameter values might be accumulated through regionalisation processes, resulting in influences on the identification of relationships between the calibrated parameter values and the catchment characteristics as well as runoff predictions in ungauged catchments.

Complex models are also often criticized for over-parameterization. It is known to be statistically unsound to model hydrographs with more than about five model parameters, where major validation of catchment models is through comparison of observed and modelled streamflow (Beven, 1989;Jakeman and Hornberger, 1993). There is widespread consensus that simpler modelling approaches, using as few parameters as possible to represent the key identifiable catchment runoff response, is a promising strategy in rainfall-runoff modelling (Nash and Sutcliffe, 1970;Beven, 1989;Jakeman et al., 1990;Jakeman and Hornberger, 1993;Young and Beven, 1994). However, an appropriate model structure based on level of complexity is still quite debatable depending on modelling objective and personal preference etc. In this paper the author therefore proposes an advanced calibration method and tests its potential for improving runoff simulations in ungauged catchments without engaging in the philosophical debate for modelling approaches, because an argument on this issue is not a focus in this study.

In addition, ultimately the choice of modelling approach depends on the research purposes and constraints, even though different levels of model complexity have their advantages and disadvantages. This study focuses on a reduction in the parameter uncertainty in regionalisation studies through improving the parameter identifiability representing the different behaviour of the catchment rather than a modification of the model structure resulting in the increase of model complexity. (i.e., modification of the model structure may potentially reduce model deficiencies and influence on the accuracy of regional models, but this is not a focus of this study.) Deficiencies on model structure suitability were identified through analyzing the variability of hydrologic response in model performance and parameter values and the NIREs (non-parametric impulse response estimates). These techniques were adopted to distinguish model parameters causing model limitations. An advanced calibration strategy was established based on the identification of deficiencies on certain parameters in representing the different flow regime. In order to give a clear meaning on the study objectives, the study focus on regionalization approach based on an advanced calibration technique will be discussed more in Section 1 (1 Introduction) in the next draft.

3. **(RC)** P7061, L21-22: Again, why not address the model structure?

(Author Comment for Specific Comments, AC 3)

Please see the Author Comment 2 (AC 2) above.

4. **(RC)** P7061, L22-25: How portable/reproducible is this to other settings? How intensive is the process?

(Author Comment for Specific Comments, AC 4)

This approach can be extended to other rainfall-runoff models using the following steps. First, the streamflow time series are isolated by a certain criterion. The classification criterion can vary with researcher's perspective such as month, season (Kim and Lee, 2014a), flow regime (Boyle et al., 2000) and flow event classified certain thresholds (Andrews et al., 2011) etc. After choosing a partitioning criterion, it is necessary to identify model parameters largely affected by the selected criterion. Then, calibration procedure adapted to a specific model structure can be established according to the identification of certain parameters in representing the selected partitioning criterion.

5. **(RC)** P7062, L13-14: Is this a big enough sample size to draw "real" conclusions from?

(Author Comment for Specific Comments, AC 5)

The possible improvements for the dry period in regionalization were identified and evaluated using 11 catchments located in Korea (the data limitation caused by significant reservoir regulation was introduced in 'Section 2. Study site and data set' of the manuscript; limited data availability for natural catchment conditions was also described in the third paragraph in AC 6 below).

The overall goal of this research is not to derive 'accurate and reliable' regional model equations, which is unlikely to be robust with the small number of catchments used, but to investigate the effectiveness of the regional models derived from the refined calibration approach through comparison with those from the traditional approach. This study focuses on identifying the differences in the relationships between the catchment attributes and the model parameters (calibrated based on the traditional and the refined approaches) and evaluating the adequacy of the regional models (from the traditional and the refined calibration approaches) based on the comparison under the same conditions, rather than evaluating the accuracy of the regional models themselves.

There may be a lake of robustness in the regional models due to the small number of catchments. This study is, however, limited to investigating the differences caused by the calibration methods through a regionalisation procedure (i.e., relationships between the catchment attributes and the model parameters in terms of a hydrological perspective and regional models etc). The results are sufficient to provide the effectiveness of the regional models from the refined calibration approach when it is compared with that of the regional models derived from the traditional calibration approach. The issue on the lake of robustness in the regional models due to the small number of catchments, which described above, will be discussed more in the last section (6. Summary and conclusion) in the next draft.

6. **(RC)** P7065, L22-25: What is the "real" impact of this? If the wet period accounts for 75% of total Q, does it really matter that dry period performance is not as good (but perhaps still tolerable)? Are stakeholders overly concerned about low flows?

(Author Comment for Specific Comments, AC 6)

The five major basins in Korea generally show a similar seasonal distribution of precipitation and runoff. More than 70% of annual precipitation falls during the monsoon season (i.e., from June to September) with peak precipitation and peak runoff occurring in July. Winter is the driest time of the year climatologically, but hydrologic drought occurs in spring when soil moisture is depleted as evapotranspiration increases. These kinds of seasonal rainfall distributions and runoff characteristics make water resource management in Korea difficult.

The major problem for water resource planning and management in Korea has been how to balance the tradeoffs between instream uses (e.g., environmental flows such as aquatic life and recreation) and out-of-stream uses (e.g., water intakes by the different water sectors such as the irrigation intake, industrial and domestic uses, and water transfer across basins). Management problems normally exacerbate during the dry period (i.e., low flow periods) with on-going water resource development resulting in gradual reduction of flow available for instream uses.

The balance between environmental flows and water intakes is determined based on information on low flow characteristics (derived from low flow estimation). Low flow characteristics during the dry period are normally based on long continuous daily streamflow time series at a site of interest. The primary source of daily data is the observed streamflow records. However, streamflow observations often cover only a short period of time, and daily time-series data for natural catchment conditions are not available for most catchments in Korea. When observed streamflow records are insufficient (or are not available in ungauged catchments), low flow characteristics may need to be estimated from simulated daily streamflow time-series for the dry periods (i.e., low flow estimation) through rainfall-runoff analysis (or regionalization).

If a model fails to estimate low flows properly (e.g., overestimation of low flows during the dry period and resulting increases in water intakes from river), a number of possible environmental effects will be caused by instream flow reduction. Such reduction may aggravate the effects of water pollution and result in changes in the density and productivity of riparian vegetation. Streamflow reduction during the dry period may also cause changes in the relative abundance of algae and organics. On the other hand if a model underestimates low flows during the dry period resulting in decreases in water intakes from river, the economic aspect (or benefits) of low flow management which is receiving increasing attention in the recent years cannot be satisfied. Some issues related to low flow management during the dry period in Korea will be added in Section 2 (2 Study site and data set) in the next draft.

7. **(RC)** P7066, L5-8: Since this is a known model structural deficiency, shouldn't the goal to be to fix the model rather than to fix the prediction?

(Author Comment for Specific Comments, AC 7)

It is well known that the water balance model calibrated over long time scales has difficulties in representing model performance during the wet and dry periods at the same time well. In this study, certain parameters (i.e., the LRM parameters in the IHACRES model) mainly causing modelling deficiency in representing the different flow regime were distinguished through analyzing the variability of hydrologic response in model performance and parameter values and the NIREs (non-parametric impulse response estimates). An advanced calibration strategy was established based on the identification of deficiencies on certain parameters in representing the different flow regime. Please see the Author Comment 2 (AC 2) for more details on an issue regarding modification of model structure.

8. **(RC)** P7066, L18-20: Because the wet period accounts for the large majority (75%) of the total Q.

(Author Comment for Specific Comments, AC 8)

The shape of the impulse response curves for the wet period was quite similar to that for the whole period in all catchments [P7066, L18-20]. The variability in the derived response curve depends on a number of factors including: the variability of the climate in the record selected; the number of sample hydrographs used to derive the response curve; as well as the catchment response characteristics etc. In this case, a high degree of similarity between the impulse response curves for the whole and the wet periods is not because the wet period accounts for the large majority of the total Q. This can be explained by the procedure for estimating the hydrologic response to an average flow event (i.e., the NIRE) adopted in this study.

An estimate of the non-parametric impulse response curve (NIRE) is a useful means of comparing hydrologic response characteristics. The method in this study used a technique for deriving a non-parametric estimate of the impulse response by obtaining an average event unit hydrograph response curve directly from observed streamflow (Croke, 2006). The average event unit hydrograph is derived solely from streamflow data (i.e., without using rainfall data).

The method used for estimating the hydrologic response to an average flow event has three steps. Identification of separated flow events having reasonably clean recessions (which are not affected significantly by previous or following flow events) in the hydrograph is the first step. This can be done by scanning for flow peaks in the stream hydrograph series and choosing peaks further apart than some minimum separation from their nearest neighbours. The selected flow peaks are then set to time step zero and the event response profiles are summed after exponential interpolation to one tenth of a time step. The peak value of the summed response curves are normalized to 1. This produces the raw response curve over an interval before and after the peaks. Finally, the procedure is then iterated to give the final profile, with subsequent flow peaks removed using the current estimated. More details are described in the research paper by Croke (2006).

As described above, flow events having reasonably clean recessions are separated. Then the event response profiles are normalized to 1 and aggregated. In this study, flow events were isolated by selecting peaks which were further apart than 10 time steps (i.e., 10 days) from neighbouring flow peaks (i.e., neighbouring peaks in the flow record within 10 days were ignored to reduce the influence of overlapping events). For the different parts of the flow regime (i.e. the entire, the wet and dry periods), a total number of the isolated flow events used in deriving the response curves in the catchments are listed in Table 1 below. The number of isolated flow events for the dry period is larger than that for the wet period. This is because the dry period (the other months) is longer than the wet period (from June to September). Another reason is that there are comparatively larger numbers of continuous peak flows, which are ignored within 10 days, during the wet period than those for the dry period. Details on the NIREs (discussed above) will be added more in Section 3 and 5.1(3 Modelling of the gauged catchment and 5.1 Assessment of the calibrated models for the entire flow regime) in a next draft.

Table 1. The number of isolated flow events in the catchments

Catchment	Number of isolated flow events for the whole period	Number of isolated flow events for the wet period	Number of isolated flow events for the dry period
Andong	76	39	51
Goesan	57	24	39
Seomjin River	94	40	74

The reason for the similar shape of the impulse response curves for the whole and the wet periods is explained through the conceptual graphs below (Figure 1). In general, hydrological response characteristics in Catchment type A (Figure 1(a)) are represented by relatively slow recession from the peak of the impulse response and strong slow flow components. These kinds of catchments may be easily calibrated over long time scale (including the wet and dry periods) due to mild recessions from the peaks and following strong baseflow components in low flows. The modelled flows are adequately calibrated between high and low flows, even though there are still marginal flow-dependent biases (i.e. underestimating and overestimating high and low flows for the wet and dry periods) in the runoff simulations.

The reason for the similar shape of the impulse response curves for the whole and the wet periods can be explained more clearly when variability in the normalized event flows for the wet period in Figure 1(a) is compared to that in Figure 1(b). The only difference between Figures 1(a) and 1(b) is a narrower range in the normalized event profiles for the wet period in Figure 1(b). The narrow range in the response profiles indicates that sharper peaks for large rainfall events and steeper recession from flow peaks are much more dominant for the wet period (compared to the wet period in Figure 1(a)) and the shapes of the response profiles are quite similar throughout the all wet periods. The reduced variability in the range of event profiles for the wet period makes the response profile for the whole period close to that for the wet period in Figure 1(b). Under these conditions, there is possibility that a tendency towards the behaviour of overestimating low flows for the dry periods increases when the model is calibrated for the wet and dry periods at the same time.

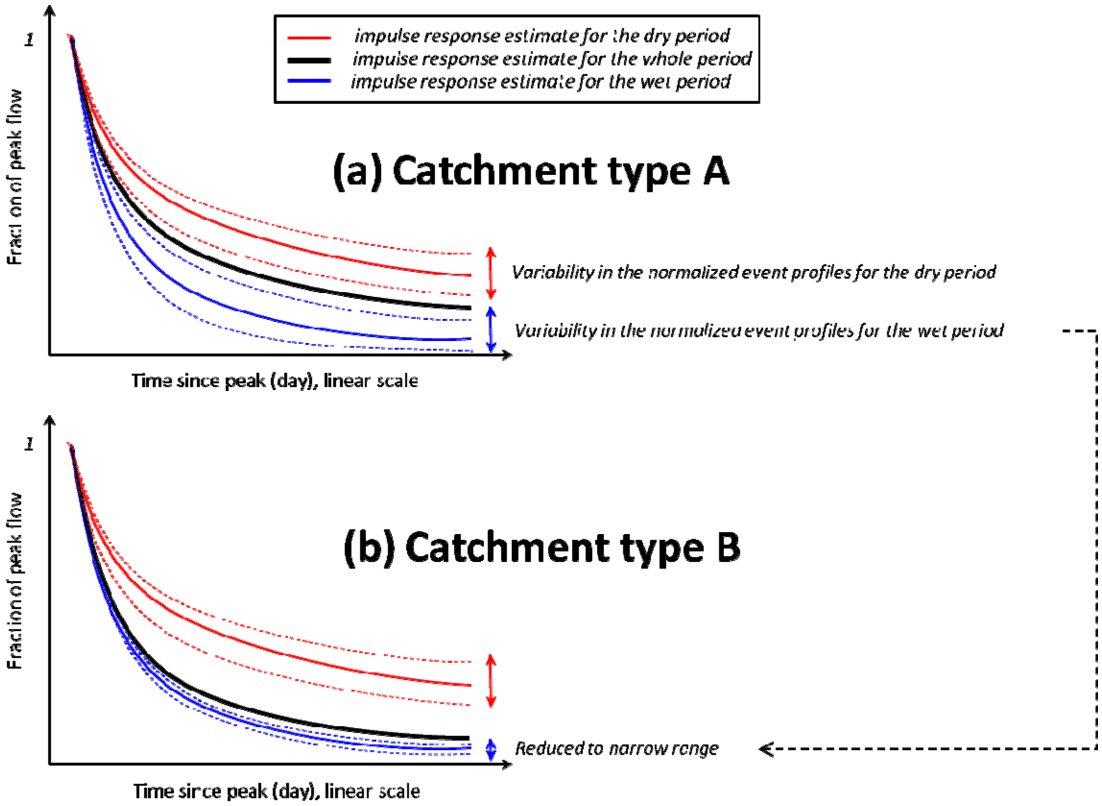

Figure 1. Conceptual graphs to explain the behavioural similarity between impulse response estimates for the whole and the wet periods

9. **(RC)** P7067, L1: I agree, although it is also connected to the larger optimization problem. Since the wet period is responsible for most of Q, the model is predisposed to do well in those times due to the optimization problem. Because of this, can we truly state with confidence that the model is "correct" during the wet period and "incorrect" during the dry period, or could the model be perhaps "incorrect" during all periods but simply optimized to the dominant period?

(Author Comment for Specific Comments, AC 9)

The author agree with the reviewer comment regarding optimization issues. The sentence [P7066 L27~P7067, L1] is also not specific. The following sentence will be removed in the next draft: [P7066 L27~P7067, L1] " The water balance model calibrated to represent the rainfall runoff characteristics over long time scales had difficulties calibrating the wet and dry periods at the same time well, suggesting deficiencies in the model structure".

10. **(RC)** P7067, L1-5: What is the real-world importance of this shortcoming? Why does this matter?

(Author Comment for Specific Comments, AC 10)

Please see the Author Comment 8 (AC 8), particularly the fifth and the sixth paragraphs. Also, please see the Author Comment 6 (AC 6) for low flow estimation.

11. **(RC)** P7067, L6-13: Again, isn't this just a consequence of the wet period having the majority of the total Q?

(Author Comment for Specific Comments, AC 11)

This is identified through comparison of the NIREs for the whole, the wet and dry periods. For more details, please see the Author Comment 8 (AC 8), particularly from the first to the fourth paragraphs of the AC 8.

12. **(RC)** P7072, L7-11: Isn't this true because the the dry period has little effect (i.e., little importance to the annual streamflow)? The results are set up to show that performance is improved at low flows (you've flipped the calibration to be based solely on this period). But despite this, the impact on the overall hydrograph is minimal. So, what is the value of doing this?

(Author Comment for Specific Comments, AC 12)

Even if the proportion of streamflow for the dry period to total flow for the whole period is relatively small, low flow estimation through an improved calibration approach is a critical issue on water resource management in Korea, particularly in ungauged catchments, during the dry period (i.e., low flow periods) with on-going water resource development resulting in gradual reduction of flow available for instream uses. Please see the Author Comment 6 (AC 6) for more details on the issues related to low flow management during the dry period in Korea. In addition, please see the Author Comment 2 (AC 2) for more details on the issue regarding a choice of modelling approach (i.e., improvement of calibration method and modification of model structure).

13. **(RC)** P7072, L21: They are reduced, but they still overlap in your box plots. Are they significantly different? Use a different word to better convey the intent.

(Author Comment for Specific Comments, AC 13)

The word, "reduced", will be revised as "improved" in the next draft. Also, the words, "the model performance variability", will be changed to "the model performance".

14. **(RC)** P7073, L1: This is, admittedly, by design.

(Author Comment for Specific Comments, AC 14)

A flow duration curve (FDC) is one of the most informative method to display the complete range of river discharges from low flows to flood events. FDCs are widely used in hydrological practice: engineering practice, low flow management and water quality management etc (Vogel and Fennessey, 1995). In the FDC analysis, various low-flow indices may be estimated from different parts of the FDC. The flows within the range of 70~99% probability of exceedance are most widely used to design low flows (Smakhtin, 2001). In this study, FDCs are plotted on linear-log and log-linear scales to better represent the high and low flows during the dry period. Particularly, FDCs on a log-linear scale are used to display low flow behaviour more clearly. In the shape of the FDC for the observed and modelled flows, the RMPS II (regional model from parameter sets II (linear routing module parameters calibrated using the refined approach over the dry period)) generally produced clearly better model performance in the low flow portion (above 70% probability of exceedance in the FDC at the log-linear scale) compared to that from the RMPS I (regional model from parameter sets I (linear routing module parameters calibrated using the traditional approach over the entire calibration period))in Figure 12 in the manuscript.

15. **(RC)** P7073, L23-26: So, what does this say about the method? Did it do what it was intended to do? It worked well at one site and not as well at another? Is your sample size large enough to draw any conclusions about the population? It seems not.

(Author Comment for Specific Comments, AC 15)

The regional models derived from the relationships between the catchment attributes and the calibrated parameter values with the refined approach over the dry period (i.e., RMPS II) and the traditional approach over the entire calibration period (i.e., RMPS I) were validated from 2003 to 2010 for the Nam River and the Seomjin River catchments. The results are listed in Table 7 in the manuscript. In addition, the model performance from the calibrated parameter sets from the refined approach over the dry period (i.e., parameter set II) and the traditional approach over the entire period (i.e., parameter set I) are listed in Table 7 in the manuscript. This is applied to identify differences in the response between the calibrated and the regional models. In Table 7 in the manuscript, the upper and lower numbers represent the model performance from the calibrated (i.e., parameter set I and II) and regional models (i.e., RMPS I and II) respectively.

In comparison of model performance between the RMPS I and II, the RMPS II definitely yielded higher model performance in *NSE*, NSE_{ln} and *%bias* at the Seomjin River and the Nam River catchments (see Table 7 in the manuscript). The sentence [P7073, L23-26], "The degraded performance with NSE_{ln} and *%bias* ...", was used to explain the differences in the response between the calibrated (parameter set II) and the regional (RMPS II) models. The model performances from parameter set II come from local model calibration. On the other hand the model performances from the RMPS II come from regionalized parameters (i.e., parameters estimated from regional models). This is not surprising that the model performances from the local calibration model are better than those from the regional model. In order to give clear meaning on the comparison between the local and regional models in Table 7 in the manuscript, additional explanation discussed above will be added in the next draft. Please see the Author Comment 5 (AC 5) for more details on additional issue related to a sample size.

16. **(RC)** P7073, L27: Due to its lack of relative importance to the overall system dynamics (as % of total Q).

(Author Comment for Specific Comments, AC 16)

It seems that the Referee Comment 16 (RC 16) is equal to the RC 17 in terms of meaning, and the RC 17 is more clear. Therefore, please see the Author Comment 17 (AC 17) for this question.

17. **(RC)** P7073, L27-29: I would dispute this. I agree that this may be true visually for low flows, but it is not necessarily a "real" conclusion relative to total Q. And, is this improved model performance (hydrologic behavior) real or simply manufactured? I think the larger problem is that the focus is on fixing the outcome rather than fixing the model structure.

(Author Comment for Specific Comments, AC 17)

Water resource management in Korea consists of flood control for the wet period and low flow management for the dry period. Each management is equally important, even if the proportion of streamflow for the dry period to total flow for the whole period is relatively small. In particular, low flow estimation is a critical issue in ungauged catchments. Please see the Author Comment 6 (AC 6) for more details on the issues related to low flow management during the dry period in Korea. Also, please see the

Author Comment 2 (AC 2) for more details on the issue regarding a choice of modelling approach (i.e., improvement of calibration method and modification of model structure).

18. **(RC)** P7076, L3-4: Due to the way the calibration was constrained - essentially the author increased the weight of the low flow period in the parameter optimization causing better performance there at the expense of the high flows.

(Author Comment for Specific Comments, AC 18)

This study investigates the possibility to improve the hydrological model parameter regionalization approach through rainfall-runoff model calibration for the different parts of flow regime, particularly focusing on low flow behavior during the dry period. In local model calibration, the parameters calibrated for the whole period were used in the wet period because of the similar dynamics in the hydrological responses (i.e. impulse response estimates) and model performance for the whole and the wet periods. The model predictions were applied to the split historical time series into the wet and dry periods, and the model was calibrated on data segments only for the dry period separately. Therefore, there was no constrain causing better performance in low flows during the dry period at the expense of improvement of high (or large) flow during the wet period (because the calibrated parameters for the whole period were directly used in the wet period). This strategy was established according to an identification of deficiencies on model structure suitability for the different flow regime. This was identified through analysis for the consistency of hydrologic response in model performance and parameter values and the NIREs. For more details on the establishment of calibration procedure, please see Section 5.1 and 5.2 in the manuscript.

TECHNICAL CORRECTIONS

(RC) P7058, L9: change “was” to “were”
P7061, L3: change “4-season” to “four seasons”
P7061, L7: change “4-season” to “season”
P7061, L15: change “was” to “were”
P7062, L7-8: change “for various reservoirs” to “of the reservoirs”
P7062, L16: change “4” to “four”
P7067, L1: delete “well”

(Author Comment for Technical Corrections)

The author thanks for the thoughtful corrections. The typos will be corrected in the next draft.

References

Andrews, F. T., Croke, B. F. W., and Jakeman, A. J.: An open software environment for hydrological model assessment and development, Environmental Modelling and Software, 26, 1171-1185, 2011.

Beven, K.: Changing ideas in hydrology -- The case of physically-based models, Journal of Hydrology, 105, 157-172, 1989.

Beven, K., and Freer, J.: Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology, Journal of Hydrology, 249, 11-29, 2001.

Boyle, D. P., Gupta, H. V., and Sorooshian, S.: Toward improved calibration of hydrologic models: combining the strengths of manual and automatic methods, Water Resources Research, 36, 10.1029/2000wr900207, 2000.

Bulygina, N., Ballard, C., McIntyre, N., O'Donnell, G., and Wheater, H. C. W.: Integrating different types of information into hydrological model parameter estimation: Application to ungauged catchments and land use scenario analysis, Water Resources Research, 48, n/a-n/a, 10.1029/2011wr011207, 2012.

Croke, B. F. W., Merritt, W. S., and Jakeman, A. J.: A dynamic model for predicting hydrologic response to land cover changes in gauged and ungauged catchments, Journal of Hydrology, 291, 115-131, 2004.

Croke, B. F. W.: A technique for deriving an average event unit hydrograph from streamflow-only data for ephemeral quick-flow-dominant catchments, Advances in Water Resources, 29, 493-502, 2006.

Dietrich, C. R., Norton, J. P., and Jakeman, A. J.: Ill-conditioning in environmental system modeling, in Modeling Chagne in Environmental Systems, edited by A.J. Jakeman, M.B. Beck, and M.J. McAleer, John Wiley, New York, 1993.

Dunn, S. M.: Imposing constraints on parameter values of a conceptual hydrological model using baseflow response, Hydrology and Earth System Sciences, 3, 271-284, 1999.

Hundecha, Y., and Bardossy, A.: Modeling of the effect of land use changes on the runoff generation of a river basin through parameter regionalization of a watershed model, Journal of Hydrology, 292, 281-295, 2004.

Hundecha, Y., Ouarda, T. B. M. J., and Bardossy, A.: Regional estimation of parameters of a rainfall-runoff model at ungauged watersheds using the spatial structures of the parameters within a canonical physiographic-climatic space, Water Resources Research, 44, W01427, 10.1029/2006wr005439, 2008.

Jakeman, A. J., Littlewood, I. G., and Whitehead, P. G.: Computation of the instantaneous unit hydrograph and identifiable component flows with application to two small upland catchments, Journal of Hydrology, 117, 275-300, 1990.

Jakeman, A. J., and Hornberger, G. M.: How much complexity is warranted in a rainfall-runoff model?, Water Resources Research, 29, 2637-2649, 1993.

Jakeman, A. J., and Littlewood, I. G.: An assessment of the dynamic response characteristics of streamflow in the Balquhidder catchments, Journal of Hydrology, 145, 337-355, 1993.

Kay, A. L., Jones, D. A., Crooks, S. M., Calver, A., and Reynard, N. S.: A comparison of three approaches to spatial generalization of rainfall–runoff models, Hydrological Processes, 20, 3953-3973, 10.1002/hyp.6550, 2006.

Kim, H. S., Croke, B. F. W., and Jakeman, A. J.: Model-based analysis to identify the impacts of climate variability and land use changes on streamflow, American Geophysical Union, Fall Meeting 2008, San Francisco, USA, 2008,

Kim, H. S., and Lee, S.: Assessment of a seasonal calibration technique using multiple objectives in rainfall–runoff analysis, Hydrological Processes, 28, 2159-2173, 10.1002/hyp.9785, 2014a.

Kim, H. S., and Lee, S.: Assessment of the adequacy of the regional relationships between catchment attributes and catchment response dynamics, calibrated by a multi-objective approach, Hydrological Processes, 28, 4023-4041, 10.1002/hyp.9942, 2014b.

Klemes, V.: Operational testing of hydrological simulation models, Hydrological Sciences Journal 31, 13-24, 1986.

Koivusalo, H., and Kokkonen, T.: Modelling runoff generation in a forested catchment in southern Finland, Hydrological Processes, 17, 313-328, 2003.

Kokkonen, T. S., Jakeman, A. J., Young, P. C., and Koivusalo, H. J.: Predicting daily flows in ungauged catchments: model regionalization from catchment descriptors at the Coweeta Hydrologic Laboratory, North Carolina, in, John Wiley & Sons, Ltd., 2219-2238, 2003.

Kumar, R., Samaniego, L., and Attinger, S.: Implications of distributed hydrologic model parameterization on water fluxes at multiple scales and locations, Water Resources Research, 49, 360-379, 10.1029/2012wr012195, 2013.

Lamb, R., Crewett, J., and Calver, A.: Relating hydrological model parameters and catchment properties to estimate flood frequencies from simulated river flows, BHS 7th National Hydrology Symposium, Newcastle-upon-Tyne, U. K., 2000.

Letcher, R. A., Schreider, S. Y., Jakeman, A. J., Neal, B. P., and Nathan, R. J.: Methods for the analysis of trends in streamflow response due to changes in catchment condition, Environmetrics, 12, 613-630, 2001.

Littlewood, I. G.: Improved unit hydrograph characterisation of the daily flow regime (including low flows) for the River Teifi, Wales: towards better rainfall-streamflow models for regionalisation, Hydrology and Earth System Sciences, 6, 899-911, 10.5194/hess-6-899-2002, 2002.

McIntyre, N., Lee, H., Wheater, H., Young, A., and Wagener, T.: Ensemble predictions of runoff in ungauged catchments, Water Resources Research, 41, W12434, 10.1029/2005wr004289, 2005.

Merz, R., and Blöschl, G.: Regionalisation of catchment model parameters, Journal of Hydrology, 287, 95-123, 2004.

Nash, J. E., and Sutcliffe, J. V.: River flow forecasting through conceptual models part I - A discussion of principles, Journal of Hydrology, 10, 282-290, 1970.

Oudin, L., Andréassian, V., Perrin, C., Michel, C., and Le Moine, N.: Spatial proximity, physical similarity, regression and ungaged catchments: A comparison of regionalization approaches based on 913 French catchments, Water Resources Research, 44, W03413, 10.1029/2007wr006240, 2008.

Parajka, J., Merz, R., and Blöschl, G.: A comparison of regionalisation methods for catchment model parameters, Hydrology and Earth System Sciences Discussions, 2, 509-542, 10.5194/hessd-2-509-2005, 2005.

Parajka, J., Blöschl, G., and Merz, R.: Regional calibration of catchment models: Potential for ungauged catchments, Water Resources Research, 43, W06406, 10.1029/2006wr005271, 2007.

Post, D. A., Jakeman, A. J., Littlewood, I. G., Whitehead, P. G., and Jayasuriya, M. D. A.: Modelling land-cover-induced variations in hydrologic response: Picaninny Creek, Victoria, Ecological Modelling, 86, 177-182, 1996.

Post, D. A., and Jakeman, A. J.: Predicting the daily streamflow of ungauged catchments in S.E. Australia by regionalising the parameters of a lumped conceptual rainfall-runoff model, Ecological Modelling, 123, 91-104, 1999.

Samaniego, L., Kumar, R., and Attinger, S. C. W.: Multiscale parameter regionalization of a grid-based hydrologic model at the mesoscale, Water Resources Research, 46, n/a-n/a, 10.1029/2008wr007327, 2010.

Schreider, S. Y., Jakeman, A. J., Falkland, A., and Knee, R.: Streamflow prediction for the Queanbeyan River at Tinderry, Australia, Environment International, 21, 545-550, 1995.

Schreider, S. Y., Whetton, P. H., Jakeman, A. J., and Pittock, A. B.: Runoff modelling for snow-affected catchments in the Australian alpine region, eastern Victoria, Journal of Hydrology, 200, 1-23, 1997.

Sefton, C. E. M., and Howarth, S. M.: Relationships between dynamic response characteristics and physical descriptors of catchments in England and Wales, Journal of Hydrology, 211, 1-16, 1998.

Seibert, J., and McDonnell, J. J.: On the dialog between experimentalist and modeler in catchment hydrology: Use of soft data for multicriteria model calibration, Water Resources Research, 38, Art. no. -1241, 2002.

Smakhtin, V. U.: Low flow hydrology: a review, Journal of Hydrology, 240, 147-186, http://dx.doi.org/10.1016/S0022-1694(00)00340-1, 2001.

Smith, T., Marshall, L., and Sharma, A.: Predicting hydrologic response through a hierarchical catchment knowledgebase: A Bayes empirical Bayes approach, Water Resources Research, 50, 1189-1204, 10.1002/2013wr015079, 2013.

Uhlenbrook, S., and Sieber, A.: On the value of experimental data to reduce the prediction uncertainty of a process-oriented catchment model, Environmental Modelling and Software, 20, 19-32, 2005.

Vogel, R. M., and Fennessey, N. M.: Flow duration curves II: a review of applications in water resources planning, JAWRA Journal of the American Water Resources Association, 31, 1029-1039, 10.1111/j.1752-1688.1995.tb03419.x, 1995.

Wagener, T., and Wheater, H. S.: Parameter estimation and regionalization for continuous rainfall-runoff models including uncertainty, Journal of Hydrology, 320, 132-154, 2006.

Wale, A., Rientjes, T. H. M., Gieske, A. S. M., and Getachew, H. A.: Ungauged catchment contributions to Lake Tana's water balance, Hydrological Processes, 23, 3682-3693, 10.1002/hyp.7284, 2009.

Yadav, M., Wagener, T., and Gupta, H.: Regionalization of constraints on expected watershed response behavior for improved predictions in ungauged basins, Advances in Water Resources, 30, 1756-1774, 2007.

Ye, W., Bates, B. C., Viney, N. R., Sivapalan, M., and Jakeman, A. J.: Performance of conceptual rainfall-runoff models in low-yielding ephemeral catchment., Water Resources Research, 33, 153-166, 1997.

Young, A. R.: Stream flow simulation within UK ungauged catchments using a daily rainfall-runoff model, Journal of Hydrology, 320, 155-172, 2006.

Young, P. C., and Beven, K. J.: Data-based mechanistic modelling and the rainfall-flow non-linearity, Environmetrics, 5, 335-363, 1994.

Zhang, Y., and Chiew, F. H. S.: Relative merits of different methods for runoff predictions in ungauged catchments, Water Resources Research, 45, W07412, 10.1029/2008wr007504, 2009.