Dear Editor,

We appreciate the precise and valuable comments from the two reviewers for improving the manuscript. The reviewers' comments have been carefully checked, and complete refocusing and rewriting of some parts of paper has been carried out. The reviewers' comments are fully addressed below.

Referee #1

GENERAL COMMENTS

The presented article is a systematically performed study on modelling runoff from a small catchment adjacent to a road. The value / novelty of the article mainly is i) in concentrating on winter time conditions that often are neglected even in studies in cold climate regions, and ii) in using and comparing the performance of four structurally different hydrological models. The main concerns related to the overall quality of the article are i) the difficulty to picture runoff generation in the study area, and thereby the difficulty to assess the results and the presented analysis on the performance of the different models, ii) some inconsistencies related to the calibration and evaluation of the models: after carefully reading the article it still is not completely evident what data was used for what and why, and iii) no references to preferential flow even though it often plays a crucial role in runoff generation in Scandinavian soils. The concerns are elaborated in the following in the "Specific comments".

R. We appreciate the referee's comments and suggestions for improving the quality of the article.

i) As regards the difficulty in picturing and accessing runoff from the Skuterud catchment by the four different models, a figure of the catchment with a picture of the outlet (Figure 1) have now been added and a more detailed explanation has been included (Page 7).

ii) For further clarification about the calibration and validation periods in the four models, an additional table (Table 3) and text have now been included (Page 17).

iii) With respect to preferential flow, this process is defined in both CoupModel and HBV. We appreciate the references provided by referee 1. These references have been reviewed and included in the text and more specifically in the Discussion (Page 30 and 31 Line 722-734).

All the referee's comments have been considered and are explained further in the specific comments section below. Modification and revision of the manuscript based on the referee's comments have been conducted as follows.

SPECIFIC COMMENTS

1) Purpose and objectives of the presented study are clearly presented and explained. However on p. 5124, ll. 24-26: How do the hydrological events (the word period used later for the three different datasets) "snowmelt" and "partially frozen soil" differ from each other? Please insert a short description to clarify what is meant by these events / periods in this study. The soil supposedly is at least partially frozen during snowmelt, so is the difference that during the event "partially frozen soil" there is no snow? There is some explanation to

this from p. 5140 on, though a short explanation already in the Introduction would clarify, what actually is investigated with the models.

R. For further clarification, the relevant paragraph has now been changed to:

"Specific aims were to:

- Identify the capability and usefulness of models of differing degrees of complexity in simulating discharge situations
- Compare the responses of the selected models during three periods. Period I represents hydrological condition when it rain or snowmelt or both occur on partially frozen soils. So this period accounts for the presence of snow on the ground and its melting and subsequent runoff and infiltration of water into the soil. This period consisted of two main events, on 10 January and 13-18 January 2008. Period II included frozen soil and rain event conditions. This period included two main events, on 24 November and 1 December 2007. Period III included non-frozen soil and rain event conditions. This period also consisted of two main events, on 6 and 9 November 2007"

Page 6 Line 121-128

2) Chapter 2.1, p. 5125: A map of the area, showing the land-use and soil types, topography, as well as the location of the road with its drainage constructions, would help in picturing the area and the circumstances for runoff generation, and in assessing the results and the analysis of the models.

R. Good point. A map of the area and a photograph of Skuterud catchment outlet have now been inserted as Figure 1.

Page 7 Line 152

3) A short description of how runoff is affecting the drainage construction would also be helpful: what part of the runoff is expected to enter the drainage system and by which mechanisms – only surface runoff important for the drainage system?

R. The following sentence has been inserted to clarify this point:

"In the Skuterud catchment, the subsurface drainage is a system of covered piped channels, often described as field drains, to remove excess soil water above the level of the drain pipes. Rising groundwater is expected to be drained away by the pipes when it reaches above the drainage level. The drains are installed at 10 m spacing and 80 cm depth."

Page 8 Line 165-168

4) The CoupModel apparently is the only model in this study that directly includes soil frost. Please highlight this on p. 5127, for instance on l. 4, after saying "Our search for hydrological models resulted in the selection of four models that met the stated criteria" – one of the criteria being "the models had to be applicable to catchments subject to winter conditions with frost, snow and frozen soil" (cf. p. 5123 ll. 25-27). The other three models include the effect of winter conditions by snow and snowmelt, or by changing parameter values (like the K-sat value) so that infiltration and unsaturated flow is delayed to include the effect of frozen soil.

R. The paragraph in question has now been changed to

"Our search for hydrological models resulted in the selection of four models that met the stated criteria. These models, which differed greatly in structure and input requirements, were LISEM, MIKE SHE, CoupModel and HBV.

The CoupModel is the only model in this study that directly includes soil frost. The other models indirectly approximate the effect of winter conditions by including snow and snowmelt statuses by changing parameter values (such as the Ks value), so that infiltration and unsaturated flow are delayed to include the effect of frozen soil."

Page 9 Line 199-205

5) Chapter 2.2.3, pp. 5129-5130: The CoupModel is presented in less detail than LISEM and MIKE SHE. Please insert at least a notion of dimensionality of the CoupModel (1D?) and a short explanation on how the flow is directed in the model to generate runoff.

R. The following text has been added with reference to the CoupModel:

"In CoupModel the vertical movement of water for each soil layer in a 1D soil profile has been defined by the Richards equation, using a water retention function (Brooks and Corey, 1964) and an unsaturated hydraulic conductivity function (Mualem, 1976). The model uses the first term in the Hooghoudt drainage equation (Hooghoudt, 1940) to calculate horizontal outflows from the saturated layer to a hypothetical drainage pipe."

Page 12 Line 279-284

6) Considering the order of Chapters 2.2 and 2.3: Table 1 listing the main features of the models is very informative. While reading through Chapter 2.2, many questions arise that are then answered when reaching section 2.3 and a reference to Table 1. I was, for instance, asking that which models do describe soil frost directly: There was only a statement about MIKE SHE in Chapter 2.2.2 that it does not describe it. Therefore, I recommend moving the reference to Table 1 right before Chapters 2.2.1-2.2.4. Chapter 2.3 could actually be removed and the few sentences in it could be moved into Chapter 2.2.

R. The reference to Table 1 and a few sentences from section 2.3 have been moved in accordance with this suggestion.

"The main features of the four models are summarised in Table 1, together with the hydrological processes examined in the comparison of these models in this study. The models also differ in complexity and in their differentiation of different flow processes such as surface-subsurface and groundwater runoff, as listed in Table1."

Page 9 and10 Line 211-215

7) Generally referring to all the four models presented in Chapter 2.2, and in relation to the modelled events: how do the different models describe preferential flow? At least the CoupModel includes a simple by-pass routine for preferential flow that can by-pass the soil matrix. Depending on the location of the different soil and land use types in relation to the catchment outlet / the road, vertical and lateral preferential flow as well as subsurface

stormflow may be significant contributors to the generation of the flow peak, in addition to overland flow. There is a bunch of studies on modelling preferential flow in Nordic conditions (e.g. Espeby 1989, Jansson et al. 2005, Gärdenäs et al. 2006, Laine-Kaulio 2011, Warsta 2011; see the end of the review). If you are certain that preferential flow can now be ignored in the analysis, please include a short statement, why, in the Introduction and/or the Methods section. In the opposite case, I'd recommend shortly discussing the effect of preferential flow on runoff generation and the model outcomes throughout the article.

R. The following text has now been added in the Discussion section:

"Contribution of preferential flow as well as subsurface flow has long been recognised and investigated (Whipkey, 1967; Stähli et al. 1996; Uchida et al, 2002; Uhlenbrook et al. 2002; Jansson et al. 2005; Laine-Kaulio 2011, Warsta 2011). Espeby (1990) and Sidle et al. (2000) stated that subsurface flow and preferential flow contributed more to the response of runoff than overland flow in steep catchment and from the slope.

The Skuterud catchment consists mainly of clayey soils, with some coarser soils in the highest areas. Conceptually the vertical construction of the soil is dense clay below the plough layer, which contributes little during storm events. With respect to preferential flow mechanisms the top soil, plough layer, can give rise to faster flow paths i) through cracks increasing vertical velocities and ii) through agricultural drain systems increasing horizontal subsurface velocities. The drains are already included in the models, apart from in LISEM, while the cracking is considered to be rather uncommon in this area, hence implying that effective parameters of hydraulic conductivity and porosity are sufficient for our study.

In this study, the CoupModel was the only model including an approach accounting for bypass flow (a model representation of preferential flow) of the soil matrix system (Haugen et al. 1992; Johnsson and Lundin 1991; Jansson and Gustafson 1987). However after calibration of the parameters, it was concluded that parameters representing preferential flow for this area were not important for the generation of rapid response of runoff."

Page 30 and 31 Line 722-739

8) P. 5133 ll. 24-25: What were the runoff generation mechanisms? Insert a short description. If they are explained later, consider inserting here a reference to that chapter.

R. This is further described in our Results section based on the results obtained. The sentence referred to has been deleted from the Materials & Methods section.

9) P. 5134 ll. 6-8: In addition to the lack of spatially distributed, measured parameter values, an even bigger problem often is the lack of measurement methods that describe the parameters reliably in the same scale as the model needs them. For example water retention data (-> unsat. hydraulic conductivity) as well as saturated hydraulic conductivity often are determined from soil core samples that produce an estimate way too small compared to the conductivities of larger scales (meters to tens of meters). One reason behind this is that the soil core samples do not include the effect of preferential flow and formation of subsurface stormflow on the conductivity estimate. The effect of cracks and fissures (especially for claye fields) as well as root holes, soil pipes formed by soil fauna and erosive action and stone surfaces (especially for forested moraine soils), affect both the parameter values and the runoff generation phenomenon vs. the model structure required.

R. This suggestion from the referee has been used and the following text has been added in the manuscript:

"An adequate hydrological description of water flow relies heavily on soil water retention, unsaturated hydraulic conductivity data and saturated hydraulic conductivity of the considered spatial domain. However, these measured data are often determined from soil core samples that produce an estimate which is much too small compared with the conductivities of larger scales (metres to tens of metres). Therefore, the lack of empirical methods that describe the parameters reliably in the same scale as the model is an even greater problem than the lack of spatially distributed, measured parameter values." Page 17 Line 399-405

10) P. 5135 ll. 17-21: Are you lowering the K-sat value of the top layer to prevent water infiltrating into the frozen soil, and, thereby producing more overland flow, forcing the runoff to compose more of water flowing on the soil surface? A direct statement of this would be good.

R. It is exactly what was done. A small addition has been made to the sentence to clarify this:

"Calibration of the model results was performed on the measured peak discharge, optimising the saturated conductivity value of the frozen layer. Initially, the saturated conductivity (Ks) values as presented in Table 2 were used in the calculations. These values were lowered for the first 25 cm of the soil profile to simulate the effect of frozen soil on the hydraulic conductivity, i.e. forcing the water to flow over the surface."

Page 19 Line 438-443

11) Chapter 2.5.4: Was the CoupModel using the by-pass routine to generate subsurface, preferential flow? Were the parameters guiding the by-pass flow among the calibrated ones?

R. Yes. The AScaleSorption parameter representing by-pass flow was among the calibrated parameters in the CoupModel. This parameter represents the sorption scaling coefficient for flow into the matric pore domain. After comparing prior mean and post mean values of AScaleSorption, we found that this parameter was not sensitive in the calibration process or the generation of runoff.

12) After reading pp. 5134-38: Why were the calibration periods of the models different, e.g., why is all the data taken into account when calibrating the CoupModel with the MC procedure (p. 5136 ll. 24-25)? When later analyzing the simulations further, is the general, better performance of the CoupModel and HBV mainly a result of that that all the data were used for finding the best parameterization, whereas the parameterization of the LISEM and MIKE SHE were tuned with only one event in one period?

R. CoupModel and HBV model were initially calibrated for 6 months. This resulted in inconsistency in calibration periods between the four models (also raised by referee 2). Therefore the runoff simulations with the CoupModel and HBV model were repeated. It should be noted that CoupModel and HBV should preferably be calibrated for a longer period in order to find the best parameterisation. Calibration of these models for a single short event does not contain any information that helps the models to be good at predicting other events. However, when using the entire period in CoupModel and HBV, the models did not show big differences compared to what was presented before.

A table (Table 3) that clarifies the periods in which models were calibrated and validated has been added:

Table 3. Calibration and validation periods of discharge in the four models used

	Hydrological event	LISEM	MIKE SHE	CoupModel	HBV
Calibration period					
Period I 10-11 January 2008	Snowmelt and rain event on 10 January	Yes	Yes	Yes	Yes
Validation periods					
Period I 13-18 January 2008	Snowmelt and rain events on 13-18 January	Yes	Yes	Yes	Yes
Period II 20 November- 9 December 2007	Frozen soil and rain events on 24 November and 1 December	No	Yes	Yes	Yes
Period III: 2-12 November 2007	Non-frozen soil and rain events on 6 and 9 November	No	Yes	Yes	Yes
Simulation time		5 days	16 months	16 months	16 months

The text has been changed as follows:

"CoupModel was run for the same 16-month simulation period as MIKE SHE, i.e. from 1 January 2007 to 30 April 2008. This model was calibrated for a one-day period, the event between 10-11 January 2008 mentioned in Table 3."

Page 20 and Line 460-462

13) P. 5140 ll. 10-12: There is no validation period for the LISEM, and the model is only evaluated against the data used for the model calibration: this could be stated directly already in the beginning of the Calibration Chapter.

R. For better clarification this sentence has now been changed to:

"The model was calibrated to the event between 10 and 11 January 2008 and validated using the event between 13 and 17 January 2008 (Table 3)".

Page 18 Line 433-435

14) Pp. 5141-5143: The use of words calibration, evaluation, etc.: would it be possible to define these concepts already in the methods section. And also explicitly tell what period or what event in a period was used for what related to each of the models. Now it is slightly difficult to find this information scattered in the calibration and results chapters.

R. Please see the answers to comments 1, 12, 13, and the new Table 3 inserted in the text.

15) P. 5145 ll. 1-2: I'd suggest mitigating the trivial message somehow, only noting for instance "Despite the common differences in measured and modelled runoff values, the timing of simulated and observed in this study."

R. This comment was considered and the sentence in question has been rephrased thus:

"Despite the common differences in measured and modelled runoff values and the timing of simulated and observed events in this study, the simple HBV model gave better prediction of simulated peak discharge than the more complex models (LISEM, MIKE SHE and CoupModel)."

Page 32 and Line 779-782

16) P. 5145 ll. 4-9: Does this now mean that the goodness of fit measures in table 3 contain the whole simulation period (containing periods I, II and III), or do they describe the calibration period that is (?) the first event in the period I, or different periods for different models? Please make clear what is behind the values.

R. For better clarification, the following text has been added and the caption of Table 5 (old table 3) has been updated:

"In order to compare the discharge from the LISEM, MIKE SHE, CoupModel and HBV models for different periods, the statistics R2 and NSE were calculated. The R2 and NSE values presented in Table 5 were obtained for the best selected simulations in LISEM and MIKE SHE. These values were compared with the minimum and maximum values of R2 and NSE for all accepted runs in CoupModel and HBV."

Page 23 and Line 533-537

Table 5 caption:

"Table 5 Statistics for the four models (R2 = coefficient of determination, NSE = Nash-Sutcliffe simulation efficiency) over the calibration period (10-11 January 2008) and three validation periods (13-17 January 2008, 20 November-10 December 2007 and 2-12 November 2007)."

Page 49

17) P. 5146 Chapter 4.2: How about the effect of preferential flowpaths on infiltration? Models used do not describe this mechanism in enough detail?

R. This question was raised in an earlier comment as well. Please see the answer to comment 7, where it has been addressed.

TECHNICAL CORRECTIONS

18) P. 5122 ll. 5-7: the sentence difficult to understand. Could it be re-written in a more understandable form so that the subject of the sentence, before the predicative, is not two lines long?

R. The sentence in question has now been changed to:

"The simulated and observed discharges generated during three types of hydrological situations characteristic of winter/spring conditions causing overland flow were considered. These three hydrological situations were snowmelt, frozen or partially frozen soil and heavy rain events."

Page 3 Line 47-50

19) P. 5127 ll. 3-4: consider repeating here the criteria for selecting the models for this study (they were listed in the Introduction on p. 5123).

R. Additional text has now been inserted as below:

"There are several criteria to be considered in choosing between a variety of hydrological models: i) model availability; ii) model performance on an hourly basis; iii) availability of input data; iv) need for calibration; v) applicability of the model to catchments subject to winter conditions with frost, snow and frozen soil; and vi) previous testing or use of the models in practical applications"

Page 9 Line 194-198

20) P. 5128 ll. 1-2 and ll. 8-9: the same sentence about infiltration repeated.

R. This repeat sentence has been deleted.

21) Table 1, the independent input data for LISEM: "distributed input" repeated.

R. This has now been deleted.

22) 5133 l. 6: a dot missing between "data" and "Vegetation".

R. Inserted.

23) Table 2, referred to on p. 5133 ln 23: Model names LISEM and MIKE SHE repeated in the table. Correct the names at the lower half of table to Coup model and HBV.

R. This has now been done.

24) P. 5133 l. 26 vs. ll. 28-29: edit the sentences such that it is directly stated which event refers to which dates / periods.

R. The text has now been modified to:

"Using data from the Skuterud catchment, it was possible to analyse the runoff during three periods, each of which included different types of hydrological events: Period I: 10 January-18 January 2008 (partially frozen soil, snow melt and rain event); Period II: 20 November-10 December 2007 (frozen soil and rain event); and Period III: 2-12 November 2007 (non-frozen soil and rain event) (Table 3)"

Page 16 and 17 Line 383-387

25) P. 5135 l. 15: Reference to Fig.3: are the soil temperatures measurements or simulated with some of the models? If they were calculated with the Coup model, please mention it at least in the title of the figure. After having read 5 pages further, it is said that the values are calculated - if they however are measurements, I suggest replacing "results" with "observations".

R. As stated on Page 23 Line 542-546, "unfortunately there were no measured data on soil temperature during these three periods for the catchment studied. The simulated soil temperature data from CoupModel were used to analyse possible runoff generation mechanisms based on a range of soil temperature and soil frost conditions related to various hydrological responses."

The legend to Figure 3 has been revised to:

"Fig. 3. Calculated temperature using CoupModel, measured precipitation and measured and simulated discharge during Period I. Simulated discharge for MIKE SHE (blue line), CoupModel (grey band), HBV (yellow band) and LISEM (green line). Measured discharge: black dashed line. The event on 10 January was used for calibration and events between 13-18 January for validation".

26) P. 5135 l. 19: Table 3 should be Table 2?

R. The numbering of tables has now been changed.

27) P. 5136 l. 23: word "as" missing between words "period" and "MIKESHE".

R. Inserted

28) P. 5139, reference to Table 3: complement the title of the table to include the information that the values are for the calibration period (as mentioned on ll. 6-8), and also tell explicitly what is the calibration period. Where are the R2 and NSE values for the validation / prediction period? Consider including those values in the same table as well, even though you use the peak flow residuals as the main criteria to analyse the errors in predicted discharges (cf. p. 5142).

R. A reference to this table has been added on Page 23 Line 535. The title of this table has been revised.

Previous Table 3 is now Table 5 in the revised version of this manuscript. The R2 and NSE values for the validation / prediction period have also been added to Table 5, which has the following legend:

Table 5 Statistics for the four models (R2 = coefficient of determination, NSE = Nash-Sutcliffe simulation efficiency) over the calibration period (10-11 January 2008) and three validation periods (13-18 January 2008, 20 November-10 December 2007 and 2-12 November 2007).

		LISEM	MIKE SHE	CoupModel	HBV
Calibration period					
Period I 10-11 January 2008	R2	0,69	0,28	Min=0,79 Max=0,8	Min=0,8 Max=0,82
	NSE	0,07	-0,52	Min=0,6 Max=0,64	Min=0,6 Max=0,67
Validation periods					
Period I 13-17 January 2008	R2	0,87	0,70	Min=0,78 Max=0,85	Min=0,73 Max=0,87
	NSE	0,31	0,16	Min=0,51 Max=0,74	Min=0,06 Max=0,82
Period II 20 November-9 December 2007	R2	NA	0,74	Min=0,68 Max=0,83	Min=0,75 Max=0,9
	NSE	NA	0,45	Min= -0,05 Max=0,35	Min=0,09 Max=0,87
Period III: 2-12 December 2007	R2	NA	0,44	Min=0,39 Max=0,85	Min=0,68 Max=0,89
	NSE	NA	-0,72	Min= -4,26 Max=0,58	Min=-1,74 Max=0,86

29) Chapter 3.2: Insert subtitles similarly than in previous Chapters?

R. This has now been done.

30) P. 5140 l. 15: reference to Figs. 3, 5, 7 -> should they be 3, 5, 6?

R. This has now been corrected.

31) P. 5141 l. 11: typo in the word “hydrograph”

R. This has now been corrected.

32) P. 5141 l. 13: word “optimised” rather “calibrated”?

R. This has now been changed.

33) While reading pp. 5140-41 and looking at Fig. 3: to clarify what is meant by “period” vs. “event(s)”, and which event in period is used for calibration and which for validation / evaluation, insert a statement to the title of the figure directly telling that event on 10-12.1 is used for calibration (if this is the case) and events on 13-18.1 for evaluation.

R. The text has now been modified; see new legend to Fig. 3 under point 25.

Referee #2

Overview

In the present manuscript, the authors propose to compare predictions of peak flows in response to heterogeneous meteorological and land conditions in cold conditions.
Therefore, they apply four different rainfall-runoff models of various complexity to simulate the discharge of a small Norwegian catchment in response to snowmelt, frozen soils and heavy rainfall. As stated in the introduction the aim of the study is to find out the most suitable model structure in terms of data availability and calibration requirements to predict peak flows that may alter the road network in this area. The authors conclude that the most conceptual model is the most suitable of the cohort due to the quality of the calibration data in the studied catchment. Nevertheless, they acknowledge that a more physically based model would be preferred in less well-monitored conditions, and that one of the models lacks the representation of processes relevant to this catchment.

General comments

The present manuscript does not constitute a very innovative contribution. In fact, studies that compare several hydrological model structures, or utilise modular structures for the same purpose, have become quite frequent over the last years (Breuer et al., 2009; Clark et al., 2008; Holländer et al., 2009; Plesca et al., 2012; Reed et al., 2004; Refsgaard and Knudsen, 1996). The main differences here is that the authors mainly target peak flow events in hourly time-step simulations, whereas most of the previous studies consider long term simulations. I appreciate that the authors put their study in a real-world frame with the stated will to correctly simulate hydrological extreme events to predict their effect on road infrastructures. However, this point is forgotten throughout the manuscript and only quickly evoked in the conclusion part without addressing the problematic introduced in part 1. If some flood warning levels
exist in the area, it would also be good to assess the ability of each model to correctly predict them with hit rates / false alarm rates for example (e.g. Roulin, 2007). The authors focus too much on, and are satisfied by, correctly repredicting the magnitude and timing of the peak flows, but they do not really speak about total volumes. Total volumes may also be relevant in the frame of extreme events prevention and infrastructure design. This may be due to the model evaluation which is based only on a quadratic evaluation of the error that makes the Nash-Sutcliffe efficiency more sensitive to peak values (Legates and McCabe Jr, 1999). Introducing some bias information in the evaluation (e.g. Plesca et al., 2012) could better constrain parameter sets.

R. As stated by referee 1, the novelty of the article lies mainly: i) in concentrating on winter conditions, which are often neglected even in studies in cold climate regions; ii) in using and comparing the performance of four structurally different hydrological models; and especially iii) in increasing our understanding of different models by comparing the responses of the selected models during three hydrological periods. Period I included snowmelt, partially frozen soil and rain event conditions. This period consisted of two events, 10 January and 13-18 January 2008. Period II included frozen soil and rain events on 24 November and 1 December 2007. Period III included non-frozen soil and rain events on 6 and 9 November 2007.

Changes in ground frost and ice formation on the ground surface may also cause large increases in surface runoff during snowmelt. Therefore, using models that are able to simulate discharge during ground frost phenomena is another novel aspect of this study. As mentioned by referee 2 and also stated in the manuscript, several studies over the last year have compared hydrological models, mainly during long periods. However, the present study focused on the runoff dynamics close to roads and specific watersheds on an hourly basis during the seasonal dynamics of one selected winter with interesting phenomena.

We agree that total volumes are also relevant in the frame of extreme event prevention and that the close connection between the volumes and the generation of runoff for individual events is key for prediction of the peak rate and the timing of the peak. This is an integral part of all four models applied and compared here.

The usefulness and correctness of the GLUE methodology adopted for CoupModel and HBV is discussable.

R. The GLUE methodology is not discussed in detail as it is not the main focus in this study. We applied a Monte Carlo approach with prior uniform distributions of parameters that are constrained to an ensemble of non-rejected simulations. However, we do not use the GLUE term since we did not use a formal likelihood approach and we believe that the likelihood aspects of the GLUE methodology are outside the scope of this paper. Discussions of the criteria used to reject non-valid simulations were of the highest interest to us and we have expanded the text to reflect this.

First, these two models are calibrated over a time period (Oct-07 to Apr-08) but run over the same 16 month period as LISEM and MIKE-SHE. Do you use the remaining time period (Jan-07 to Sep-07) to spinup those models?

R. Yes, this has now been clarified in the text.

This calibration period covers the Periods I, II and III later examined more in details. A validation of any sort is lacking for these calibrated models.

R. We agree with the reviewer that the previous version of the manuscript did not describe the importance of the calibration period compared with any validation period (the inconsistency in calibration periods between the four models was also raised by referee 1). Instead, we were interested in investigating the specific performance for short periods after a general conventional calibration using a longer period. The CoupModel and HBV model were initially calibrated for 6 months. We have now extended the work by including a specific short calibration period that is similar for all models, irrespective of the validation period. A table to clarify the periods for which models were calibrated and validated has been added to the revised version of the manuscript (new Table 3).

Second, while the authors introduce a threshold corresponding to R2 > 0.6 and NSE > 0.6 to discriminate between behavioural and non-behavioural parameter sets, this aspect is skipped in the presentation of results. The advantage of having quick models probably resides in the possibility to address the predictive uncertainty, especially if the calibration strategy is based on a GLUE approach that does not aim at finding a best parameter sets. Therefore, I would expect uncertainty bounds rather than single predictions in the hydrographs.
Third, performing 1,000 Monte-Carlo runs is probably much too low to find a global optimum with 17 parameters.

R. To address this very valid point, the CoupModel and HBV models were re-run for 60,000 runs. The new criteria applied were R2 > 0.79 and NSE > 0.6 to determine accepted runs in HBV and CoupModel, using a single short event as calibration. An ensemble of accepted runs from CoupModel and HBV was illustrated against the best single run simulated by LISEM and MIKE SHE (Fig. 3, 5, 6). In order to evaluate models against each other, a table of R2 and NSE values from one calibration period and three validation periods was created (Table 5). For further evaluation of the models, the peak flow residuals in Period I (which was the calibration period and the first validation period) were updated. The median values of accepted runs for CoupModel and HBV were compared with the best single value from MIKE SHE and LISEM (Figure 4). The cumulative distribution functions (CDFs) of the residuals for CoupModel and HBV have also been updated in Figures 7, 8 and 9.

Finally, if the aim of the study is to correctly predict high flows, why not targeting the sole evaluation of these events in the uncertainty analyses? Good metrics over a 6 month period do not necessarily imply a correct representation of punctual events. It would make results more comparable with LISEM's.

R. We agree that the performance for specific events is not given by a general 6-month calibration and because of this it is of high interest to evaluate the performance for single events. We have redone the calibrations to be exactly like those of the LISEM model for a short and single event. This allows us to have a more restricted comparison with LISEM, but the calibration periods were also selected for a more general view. The general calibration for the longer period can be expected to be more robust and that for the shorter period may be better for some specific events.

Based on these comments, I do not think that the manuscript is suitable for publication in HESS. Authors will find hereafter specific comments that may help them improve the presentation of their work.

R. I hope that after considering both referees' comments and suggestions and also our substantial revision, this manuscript will now be considered suitable for publication in HESS.

Specific comments

Title

1) The title should contain the word "comparison" and give a hint on the location of the study.

R. The title has been changed to "Comparison of four hydrological models in simulating high-resolution discharge dynamics of a catchment adjacent to a road in South-East Norway"

Abstract

2) The authors state that "All four models were calibrated using hourly observed streamflow" which is not true. The last sentence of the abstract may be removed.

R. This is in fact true, as all four models were calibrated against measured streamflow at the catchment outlet for the event on 10 January 2008.

Introduction

3) A recent paper by Coumou and Rahmstorf (2012) tends to lower the influence of climate change on extreme event occurrences.

R. This paper is now referred to in the Introduction:

"One of the effects of accelerating climate change is an increase in the frequency of extreme weather events in various parts of the world (Schneider et al., 2007; Green Paper EU, 2007). However Coumou and Rahmstorf (2012) question the influence of climate change on extreme event occurrences." Page 4 Line 63-66

Material and Method

4) The Material and Method part needs a major reshaping. The input data part (2.4) should be merged with the catchment description (2.1) although some model specific information (e.g. P 5133 L10-15) should be placed along the model description. Similarly, part 2.3 should be merged with part 2.2. Since there are substantive differences from model to model, setup procedures would find a better place along model descriptions.

R. The input data section (2.4) contains information about the input data used in each model. Therefore we would prefer to keep this part after 2.2.1-2.2.4. However, in accordance with the suggestions of both referees, part 2.3 has been merged with part 2.2.

Page 9 and10

5) Generally, some more details are required about the setup of CoupModel and HBV in comparison with the extended description of LISEM and MIKE SHE: are they lumped models? Semi-distributed? etc...

R. The following additional clarification has been inserted:

"In CoupModel the vertical movement of water for each soil layer in a 1D soil profile has been defined by the Richards equation, using a water retention function (Brooks and Corey, 1964) and an unsaturated hydraulic conductivity function (Mualem, 1976). The model uses the first term in the Hooghoudt drainage equation (Hooghoudt, 1940) to calculate horizontal outflows from the saturated layer to a hypothetical drainage pipe."

Page 12 Line 279-284

“In the present study, a lumped version of the HBV soil module was coupled with modules of vegetation and snow originating from the CoupModel platform (KTH, 2012). The model is described in greater detail in previous papers (Bergström, 1992; Lindström et al., 1997; Seibert, 1997)”.

Page 14 Line 312-315

Further explanation about the setup of the CoupModel and HBV models can be found under 2.4 “Calibration and setup strategies” and also in sections 2.4.4 and 2.4.5

6) P5125 L25:
To which period does this average correspond?

R. The following sentence has been inserted to clarify this point:

“The normal annual average precipitation (1961-1990) is 785 mm, with a minimum of 35 mm in February and a maximum of 100 mm in October”.

Page 7 Line 155-156

7) P5126 L5:
Please indicate with which method PET was calculated.

R. The following text has been inserted in response to this point:

“Potential evaporation (PET) was calculated using climatological data collected at the IMT station and the Penman function to represent evaporation from an open water surface (Deelstra et al., 2010).”

Page 8 Line 162-164

8) P5126 L25: Initial conditions are still important regardless of the simulation length. The state-of the-art way to lower their influence is to use a spin-up period. This needs to be clarified here.

R. The following text has been inserted:

“Hydrological simulations are commonly initialised by driving the model repeatedly over a given period until it reaches an equilibrium state. However, this type of "spin-up" approach lowers the influence of biased initial conditions, affecting modelling results during the study period (referee 2)”

Page 9 Line 187-190

9) P 5130 L22-24:
Wrong. Please check Fig 3 p. 280 of Lindström et al. (1997): there is an exponential parameter “BETA” that is used to calculate the amount of soil water recharging the flow generation boxes even when moisture conditions are below field capacity. This makes sense; otherwise the model would only simulate saturation excess processes.

R. Thanks to referee 2 for noticing this mistake. L22-24 has been deleted and the following text inserted instead:

“The model chooses a bucket approach to represent the field capacity and thus the storage capacity of the soil. The soil moisture accounting of the HBV model is based on a modification of the bucket theory in that a power coefficient parameter ‘BETA’ is used to calculate the amount of soil water recharging the flow generation boxes even when moisture conditions are below field capacity (Lindström et al., 1997).”

Page 13 Line 305-309

10) P 5132 L16:
Why not using the same delineation than in LISEM? This would make the two models more comparable.

R. At the time of setting up the MIKE SHE model, a finer delineation was available. The model was sensitivity tested against 10 m and 1 m grid and it was found that resolution had no impact on improving the simulated discharge.

11) P 5132 L18:
Why is this depth map not used in LISEM?

R. As indicated in the text: “The input data to the MIKE SHE model included data on topography, land use, vegetation, geology, hydrogeology and meteorology for the Skuterud catchment. These data were previously used in setting up the LISEM model.” MIKE SHE uses the same depth maps as LISEM.

Page 15 Line 344-346

12) P 5134 L24:
Cite Nash and Sutcliffe (1970) here.

R. This has now been done.

13) P5135 L15:
Figure 3 is described before Figure 2.

R. This has now been corrected.

14) P5135 L19:
Do you refer to Table 2?

R. Yes. This has now been corrected.

15) P5135 L23-27:
Please give more details on the calibration procedure: is it automatic? manual? etc...

R. The following sentence has added to clarify this point: “The calibration procedure is a manual adaptation of the Ks value and the results are compared with peak amount and timing.” Page 19 Line 448-449

16) P5136 L10-23:
This should go in the Results part.

R. The sentence has now been moved to the Results.

Page 21 and 22 Line 501-511

17) P5137 L4:
Please precise the distribution and ranges.

R. A new table has been created in response to this point (Table 4 on page 20):

Table 4. List of parameters used for the CoupModel and the HBV model

All Modules	Parameter	Unit	Prior		
			Min	Max	Mean
Drainage and deep percolation	DrainSpacing	m	0	10000	5.005e+3
Soil evaporation	EquilAdjustPsi		0	5	1.5
	MaxSurfDeficit	mm	-4	0	-2
	MaxSurfExcess	mm	0	3	1.25
	MaxSoilCondens	mm/day	0	1000	2.5
	KBMinusOne	-	0	5	1.25
Soil frost	HighFlowDampC	Vol %	0.1	80	40.05
	HighFlowCondImped	-	0	10	5
Soil water flows	AScaleSorption	-	0,001	1000	5.05
Surface water	SurfCoef	-	0	1000	2.05
Snow pack	MeltCoefAirTemp	Kg/Cm2day	0	10	2.5
	MeltCoefGlobRad	Kg/J	0	3e-006	5.015e-4
	OnlyRainPrecTemp	C	-4	10	-1.5
	OnlysnowPrecTemp	C	-10	4	3
Plant	IStart Value (1)	-	0	2	1
	IEnd Value (1)	-	0	2	1
Potential Transpiration	Conduct Max (1)	m/s	0.005	0.05	0.0275
HBV	BetaCoef	-	1	6	3.5
	Critical Uptake Frac	-	0.1	1	0.55
	Discharge Alfa	-	-0.5	2	0.5
	Discharge K1	1/day	0.001	0.3	0.2
	Discharge k2	1/day	1e-005	0.1	0.0505
	FieldCapacity (FC)	mm	10	2000	205
	Intial Base Storage	mm	0	3000	150
	Intial Peak Storage	mm	0	3000	150
	Intial Soil Storage	mm	0	2000	150
	PMaxPerc	Mm/day	0	5	0.35

18) P5137 L5-10:
Thousand Monte-Carlo runs for 17 parameters is very low! There is a high risk of nonuniqueness of parameter sets. How did you choose the threshold of NSE > 0.6? Since NSE is biased toward peak values, a high threshold is probably more appropriate.

R. The number of Monte-Carlo runs has been increased to 60,000 in the revised version of this paper. As stated on page 22: "According to Moriasi et al. (2007), the calibration performance for hydrological models is considered satisfactory when NSE>0.5 and R2>0.5." In addition to this statement, since the maximum value of NSE calculated in CoupModel and HBV was 0.7, a threshold of 0.6 was chosen to identify the accepted runs."

Note that we opted to use both NSE only and NSE together with R2. It proved to be important to have both the combinations of the two indicators and the level was decided in order to obtain a sufficiently large ensemble of accepted simulations.

19) P5138 L1:
See previous comment on the number of Monte-Carlo runs and threshold used.

R. This has now been changed; see the answer to the previous comment.

20) P5138 L9-13:
This should go in the Results part with some more details (number of accepted parameter sets, etc...).

R. This has now been moved to section 3.1 (Performance of the models).

Page 22 Line 525-532

21) P5138 L14-24:
This part is not necessarily needed. Authors should mention which operating system they used.

R. The following text has been inserted:

"All computations were run on a standard personal computer (Quad CPU, 8 GB installed memory and 64-bit Microsoft Windows)".

Page 21 Line 485-486

Results

22) P5139 L2-18:

These 3 paragraphs are redundant with the methods.

R. The redundant parts have been deleted and the whole section 3.1 'Performance of the models' has been revised as follows, based on comments 16, 20, 22 and 35 from referee 2:

"**3.1 Performance of the models**

The LISEM and MIKE SHE model results were optimised for measured peak discharge and timing. The Manning's n value of the channel was used to optimise for timing in both models. The saturated conductivity (Ks) value in LISEM and the drainage time constant in MIKE SHE were altered to calculate the peak discharge.

The drainage time constant is the first order coefficient of the linear reservoir model for interflow and drains. The time constant determines the velocity of drainage and mainly has an influence on the peak of the hydrograph. The smaller the time, the smaller the peak of the hydrograph (Vásquez et al., 2002). According to DHI (1998), a typical feasible value for the drainage time constant is (1x10-7, 1x10-6) s^{-1}, which is approximately equivalent to an interval of (120, 10) days. Refsgaard (1997) suggested a drainage coefficient of 33 days after calibration of a medium-sized catchment with geology consisting of sand and gravel with few moraine clay layers. For catchments with an upper geological clayey loam layer, a higher drainage coefficient was suggested by Vásquez et al. (2002). On this basis, the drainage time constant for calibration of MIKE SHE in the Skuterud catchment was set to (5.5x10-7) s^{-1}, which was equivalent to an interval of 65 days (Table 2).

In CoupModel and HBV, 60,000 simulations were run to get the parameter range, thus obtaining a reasonable distribution of output. R2 and NSE were used to evaluate simulation performance. These coefficients were chosen as likelihood measures to evaluate the accuracy of both the magnitude and timing of predicted discharge (e.g. Andersen et al., 2001; Beven, 2001; Vásquez et al., 2002; Tague et al., 2004). To date, no absolute criteria for judging model performance have been firmly established in the literature. According to Moriasi et al. (2007), the calibration performance for hydrological models is considered satisfactory when NSE>0.5 and R2>0.5. By applying the statistical criteria NSE>0.6 and R2>0.79, the number of accepted runs in this study was 273 and 120, respectively. The posterior mean and coefficient of variation (CV) value of parameters based on the accepted runs were compared with the corresponding values from the prior assumptions, based on a uniform distribution. A narrow distribution of posterior parameter values can be considered to be reduced parameter uncertainty.

Of the calibrated parameters in CoupModel, SurfCoef from the surface water module was the most sensitive. However, of the parameters in the HBV model, Discharge Alfa and Field Capacity (FC) were the most sensitive. The SurfCoef parameter is the first order rate coefficient used when calculating the surface runoff from the surface pool exceeding the residual storage (the maximum amount that can be stored on the soil surface without causing any surface runoff) (Jansson and Karlberg, 2004). Discharge Alfa is a measure of non-linearity for fast flow from the upper response box in HBV (Lindström et al., 1997) and FC is a measure of maximum soil moisture storage capacity in the soil box (Seibert et al., 2010)

In order to compare the discharge from the LISEM, MIKE SHE, CoupModel and HBV models for different periods, the statistics R2 and NSE were calculated. The R2 and NSE values presented in Table 5 were obtained for the best selected simulations in LISEM and MIKE SHE. These values were compared with the minimum and maximum values of R2 and NSE for all accepted runs in CoupModel and HBV."

Page 21-23 Line 496-537

23) P 5140 L3-4:
Please indicate to which model these values of NSE and R2 should be attributed.

R. This part has been rewritten. Please see the last paragraph of the 'Performance of the models' section above.

Page 23 Line 533-537

24) P5140 L10-12:
This should be included in the model description. One could ask why using this model at all.

R. This line has been added to the model description: "It should be noted that in this study the sub-surface drainage process was not incorporated in the model, although it is under development". Page 11 Line 247-248

25) P5143 L4-5:
Please quantify this difference. Alhough HBV and CoupModel use the same evapotranspiration module, the calibration is realised to match runoff so this difference is not very surprising as some compensation in the water balance can occur.

R. This further explanation has now been added:

"HBV simulated actual evapotranspiration almost 50% higher than CoupModel during Period III. In the HBV model, actual evaporation and groundwater recharge from rainfall were computed in the soil routine as functions of actual water storage and maximum soil moisture storage capacity (FC) in the soil box. FC was one of the most sensitive parameters calibrated by the model, with higher values reflecting greater soil water storage capacity and therefore larger water availability for evaporation."

Page 26 Line 623-628

26) P5143 L12-14:
Please describe the acknowledged process.

R. This further explanation has been added:

"However, the large difference in response to rain in the beginning of November indicated that the evapotranspiration process may be of great importance in describing some early peak runoff when the soil is partly saturated. Subsequently, more runoff would be generated from the watershed when the evapotranspiration is lower. In this study area, the actual evapotranspiration simulated in the models had a substantial influence on runoff generation in Period III. Therefore the peak runoff simulated by CoupModel and MIKE SHE, both which had lower evapotranspiration, was higher than the peak runoff generated in the HBV model, which had higher evaporation. The CoupModel estimated the evaporation as a combination of soil evaporation and transpiration."

Page 27 Line 634-643

27) P5143 L 21-23:
Large errors in MIKE-SHE and LISEM are negative ones, i.e. due to flow underestimation. This may be in relation to the Nash-Sutcliffe evaluation of the models which is biased toward high values.

R. The calibration procedure was a manual adaptation of the Ks value and the drainage coefficient in the LISEM and MIKE SHE models. Therefore, NSE was not used as a criterion to find the best parameter to give the best result. Instead, the results from different simulations by each model were compared with the measured peak amount and timing and the best simulated discharge was chosen. The NSE and R2 values presented in Table 5 were calculated for the best selected simulations in LISEM and MIKE SHE. These values were compared with the minimum and maximum values of R2 and NSE for all accepted runs in CoupModel and HBV.

Discussion

28) The authors should place their results in the frame of previously published studies and not provide a simple summary of the results part.

R. The Discussion part has been revised in response to this comment:

"Discussion

Hydrological modelling of runoff

The comparison of models indicates that outputs from models can differ considerably from measured catchment runoff values. These results are qualitatively similar to those of Deelstra et al. (2010) done for exactly the same area. However, this modelling comparison was limited to model performance evaluation over longer periods and not hourly based. In contrast, the modelling approach used our study allowed us to discuss seasonal processes focusing on winter hydrological behaviour in more detail.

Hydrologial model as a process learning tool

Runoff simulation during snowmelt: During the spring snowmelt period, a large amount of water becomes available. As the soil infiltration capacity is limited, all the water cannot infiltrate the soil. Infiltration into soil is controlled mainly by soil porosity and the effective hydraulic conductivity determined by the distribution of liquid and frozen soil moisture. Because of differences in snow simulation (snow accumulation and melting rates) in period I (Fig. 3), CoupModel, HBV and MIKE SHE produced a slightly earlier peak than the observed discharge and compared to LISEM model. The reason might be that the simulated snowmelt of these models was too early and faster than the observed runoff. MIKE SHE also underestimated the total volume of runoff. There are two possible explanations: firstly, the total accumulated amount of estimated snow was too small, hence less total volume of water available for runoff and infiltration; secondly, partially frozen soil conditions resulted in less infiltration into the soil and greater surface runoff in reality than predicted by MIKE SHE. In period I, the snow simulation had a great impact on the performance of runoff simulation.

Runoff simulation during frozen soil conditions: The hydraulic conductivity and hence the infiltration capacity of the top soil is strongly influenced by the total water content as well as the distribution between liquid and frozen water in the soil pores (Granger et al., 1984; Johnsson and Lundin, 1991). For example, CoupModel reduces the infiltration rate based on soil temperature, ice content and soil properties, while MIKE SHE does not change the infiltration capacity even when the soil is frozen. The simulation results show that MIKE SHE and CoupModel, both differing in their frozen soil schemes, overestimated total runoff during period II (Fig. 5). This can be explained by the fact that the parameters used were determined from another event without any frozen soil. In addition a significant impact of evapotranspiration on runoff was noticed during autumn.

Contribution of preferential flow as well as subsurface flow has long been recognised and investigated (Whipkey, 1967; Stähli et al. 1996; Uchida et al, 2002; Uhlenbrook et al. 2002; Jansson et al. 2005; Laine-Kaulio 2011, Warsta 2011). Espeby (1990) and Sidle et al. (2000) stated that subsurface flow and preferential flow contributed more to the response of runoff than overland flow in steep catchment and from the slope.

The Skuterud catchment consists mainly of clayey soils, with some coarser soils in the highest areas. Conceptually the vertical construction of the soil is dense clay below the plough layer, which contributes little during storm events. With respect to preferential flow mechanisms the top soil, plough layer, can give rise to faster flow paths i) through cracks increasing vertical velocities and ii) through agricultural drain systems increasing horizontal subsurface velocities. The drains are already included in the models, apart from in LISEM, while the cracking is considered to be rather uncommon in this area, hence implying that effective parameters of hydraulic conductivity and porosity are sufficient for our study.

In this study, the CoupModel was the only model including an approach accounting for bypass flow (a model representation of preferential flow) of the soil matrix system (Haugen et al. 1992; Johnsson and Lundin 1991; Jansson and Gustafson 1987). However after calibration of the parameters, it was concluded that parameters representing preferential flow for this area were not important for the generation of rapid response of runoff.

Choice of model for practical applications

In order to simulate a particular hydrological behaviour of catchments near road structures, the choice of an appropriate model structure, identifiability of parameter values and minimisation of model analytical uncertainty are vital (Son and Sivapalan, 2007). The appropriate choice of hydrological modelling tool is determined by the type of flood, and also by the length, quality and availability of data records. The chosen model structure must be relatively simple and use parameters that can be identified either from field-measured data or from analysis of catchment response data (Son and Sivapalan, 2007). In the present context, i.e. a basin where a high-quality, real-time monitoring system is available, a simple HBV model structure calibrated for a single short event also appears most suitable for a long period. It should be noted that CoupModel and HBV should preferably be calibrated for a longer period in order to find the best parameterisation. Calibration of these models for a single short event does not contain any information that helps the models to be good at predicting other events. However, when modelling the entire period in CoupModel and HBV, big differences compared to what is presented above were not found.

For ungauged basins with no real-time monitoring of discharge, MIKE SHE can be suitable because of its model structure and less dependency on calibration procedures. For example, in this case study, considering the number of parameters involved in the simulation of the entire hydrological system, and considering the fact that the simple intuitive calibration was done, a reasonable match between the observed and the simulated hydrograph at the catchment outlet could be achieved. However, the accuracy of all models was compromised by the uncertainty in physical parameters and by model structure. MIKE SHE is a flexible modelling system that integrates surface, subsurface and groundwater flow. Due to its capability as a physically-based and fully distributed model, it can be used to evaluate the impacts of alternative land use management practices on watershed response. LISEM is a single-event, physically-based model but it is potentially capable of calculating runoff from a small catchment during winter and spring. The model version used in this study was not completely adapted to the climate region studied. Therefore, modifications to snowmelt, infiltration into frozen layers and tile drainage are on-going. A new version with a snowmelt routine and drainage discharge is being tested, and first publications will be published soon.

The present study covered winter/spring conditions only. Similar comparisons of the models should be conducted for other seasons too. Extreme weather events resulting in high flows can occur at any time of the year in Scandinavia. Those already occurring under the current climatic regime can cause considerable damage to transport infrastructure (Kalantari and Folkeson, 2012)."

Page 29-32 Line 690-774

- They should also put more emphasis on the effect of floods on roads in their catchment, the actual purpose of the modelling effort as stated in the abstract and introduction.

R. The first paragraph in the Conclusions part now places more emphasis on the actual purpose of this modelling:

"Changes in climate variables will have effects on watershed hydrological responses and thus influence the amount of runoff reaching transport infrastructures. In view of accelerating climate change, there is a great need for tools such as hydrological models to quantify these changes and assess their impacts on discharge dynamics, including peak flows. Current models used for estimating peak discharge of water ways crossing roads are often based on the rational formula, one of the simplest and most widely used methods in engineering applications (Ben Zvi, 1989; Maidment, 1993). This is also used in the Norwegian handbook for road construction (Statens vegvesen, 2011). Although physically based model are required for any meaningful prediction of discharge as a consequence of climate change, these types of models are most likely too data demanding for every single crossing point to be dimensioned. This modelling exercise is a first step on the way to suggest modifications or new strategies for future guidelines. Models used to simulate total runoff in designing road drainage structures should take account of periodic hydrological behaviour in the current climate and should also be able to model future climate scenarios. Therefore this study attempted to evaluate the performance of four different models during three different periodic hydrological behaviour with focus on winter condition."

Page 32 and 33 Line 776-791

- Are there specific water level thresholds? and how good are the models to correctly predict them?

R. For the study area no threshold values are identified. There is a maximum amount of water that can be drained from the area, limited by the design of the drainage construction under the road crossing. During the measurement period (15 years), this limit had never been reached.

29) P5145 L13:
Seventeen calibrated parameters is NOT low, especially with only 1,000 Monte-Carlo runs.

R. This sentence "Seventeen calibrated parameters …" has been deleted. The number of Monte-Carlo runs has been changed to 60,000.

30) P 5147 L11-15:
Is it a planned improvement?

R. This additional information has been added: "The version of LISEM used in this study was not completely adapted to the climate region studied. Therefore, modifications to snowmelt, infiltration into frozen layers and tile drainage are ongoing. A new version with a snowmelt routine and drainage discharge is being tested, and preliminary results will be published soon." Page 32 Line 762-765

References
31) P5129 L5: Missing reference to Kristensen and Jensen (1975). Done
32) P5131 L16-17: Missing reference to Wesseling et al. (1995). Done
33) P5132 L23: A proper reference to Van Genuchten's work is needed.

R. A complete reference to van Genuchten, M. T., 1980 has been added.

34) P5133 L4-5: Missing references to both Monteith (1965) and Allen et al. (1998). Done

33) P5133 L20: Kværno and Deelstra (2003)? Yes

Tables
34) In Table 3, why is there no NSE for LISEM. I do not think you should compare criteria between models when considering different time periods.

R. NSE was calculated for LISEM. The runoff simulations with CoupModel and HBV model were performed again. A table to clarify model calibration and validation periods has been added to the revised version of manuscript (new Table 3).

The previous Table 3 (now Table 5) shows R2 and NSE over the calibration period (10-11 January 2008) and three validation periods (13-17 January 2008, 20 November-10 December 2007 and 2-12 November 2007).

Figures
35) Figure 2 is not necessary. Consider deleting it. Deleted
36) Please remove LISEM from the captions of Figure 5 and Figure 6. Removed

I hope the responses and revisions are sufficient for the revised manuscript to be accepted.
Yours sincerely,

Zahra Kalantari