Reviewer 1

The article presents the WALRUS model, a parsimonious conceptual rainfall-runoff model. Reading this paper reminded me the time I was a student and Piet Warmerdam had kindly sent me the details of the Wageningen model that I tested and found efficient. So I am pleased to see that work on this model is still ongoing and that great progress was made over the past years.

I found the article very clearly presented. I have only a few comments below. I think the article could be published after minor revision.

Detailed comments

1. p. 1358, lines 4-5: I am not sure conceptual models were mostly developed for mountainous catchments. Or maybe mountainous is not the right word here (same comment p. 1359, lines 16 and 25).
We changed "mountainous" to "sloping".

2. p. 1359, line 6: The way the WALRUS model is presented seems to indicate that it is a simulation model rather than a forecasting model (see e.g. the discussion on terminology by Beven and Young, 2013).
We changed "forecast" into "simulate". Of course, WALRUS could also be used for forecasting purposes, but that is not treated in this paper.

3. p. 1360, line 21: values lead
We corrected the typo.

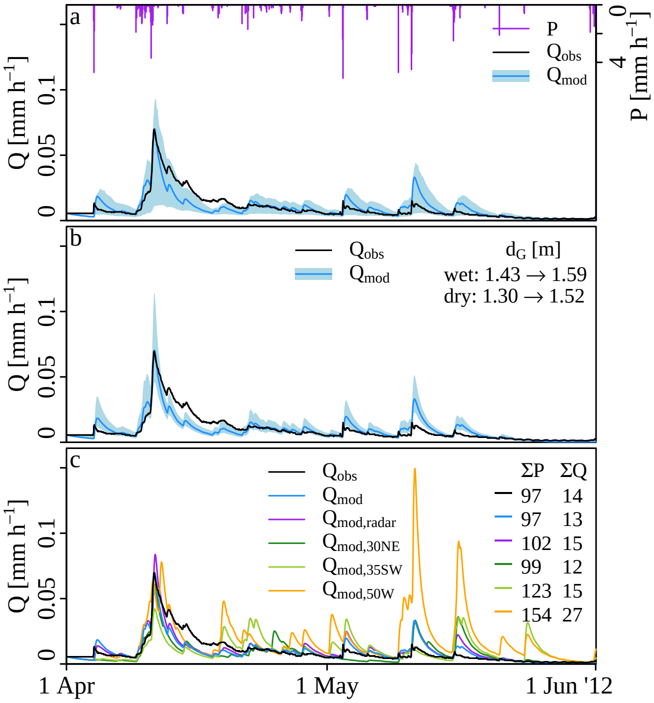
4. p. 1361, lines 24-26: This is an interesting point. Could the authors detail a bit more the limitations identified by the users on the previous model version and the way these feedbacks were collected and analysed?
We added: "Although the Wageningen Model has been widely applied in many catchments inside and outside The Netherlands, users have indicated a number of serious shortcomings, both of numeric and conceptual nature." We feel that a complete list of the shortcomings of the Wageningen Model would draw the attention from the main message: that WALRUS is a completely different model, which uses only the good parts of the Wageningen Model. A few of the practical shortcomings are: (1) too many parameters causing overparameterization and difficulties for parameter estimation, (2) long burn-in periods and (3) confusion about various versions in different programming languages.

5. p. 1362, lines 10-12: I quickly went through the companion article submitted in HESSD, which I found interesting. Since model evaluation is detailed in that paper, may I afford to make a suggestion to the authors on this other paper? I would find useful to have the performance of the new WALRUS model compared to the previous version of the Wageningen model on the two test catchments, to better quantify the improvements brought by the new modules and formulations.
WALRUS is not an improved version of the Wageningen Model, but rather a brand new model inspired by the Wageningen Model. An intercomparison study is a good idea, but we would prefer to compare it not only the the Wageningen Model, but also to other conceptual rainfall-runoff models. This would be outside the scope of the HESS paper, but we hope we can perform this research in the near future.

6. p. 1371, lines 1-15: I found this test not so useful but I leave the authors decide whether it should be kept in the paper.
We decided to keep the artificial events to test if the model produces the feedbacks we intended. In real situations, it is hard to see the effect of a certain model element, because natural variations in the forcing capture the eye.

7. p. 1378, lines 21-25: It was not fully clear for me how this function is formulated in case no observation series are available. Is it parameterized?
Seepage and surface water supply are not parameterized, because they do not depend on processes within the catchment. In case no data are available, a groundwater or surface water management model could be used to obtain time series of seepage and surface water supply, which can be used as input for WALRUS. We also added this explanation to Section 5.10.

8. p. 1382, section 5.2: Despite this initialization, it is probably still necessary to have some warm-up period. How this is managed in the software? Is there a 1-year default warm-up period? Can this be adjusted by the user? Is there any difference between long-period and event applications?
The user can decide to use a warming-up period. The option is already implemented in the code. By default, the warming-up period is set to zero, but can be changed easily by the user. We added the following sentence to Section 6.2 (5.2 in the GMDD-paper): "A user can choose to use a warming-up period in case the uncertainty around the initial conditions is large. It is implemented in the code, but by default, the warming-up period is set to zero.".
We did not need a warming-up period for the analyses we did in the HESSD-paper. Without a warming-up period, the estimate of the initial groundwater depth has a large influence on the model performance of the first few weeks, as shown in the middle panel of Figure 15 of the HESSD-paper (re-printed below), which shows the effect of uncertainty in initial conditions on discharge simulations. The effect of erroneous initial conditions on the model performance is of course larger for short periods and drier periods.



9. p. 1383, section 5.3: Could the authors explain whether there are differences in the way parameters are optimized when the model is applied on a long continuous period or on events. For example, there are parameters responsible for low flows, which may not be well-identified in case the model is calibrated on flood events only.
This is a valid remark, but we expect that this effect is limited because the number of parameters is low. In addition, the periods used for calibration did not contain extended dry periods or floods. Note that the HESSD paper (Section 5.5) contains an analysis of the effect of the objective function used for calibration and found the expected variations in resulting model parameters. This analysis already shows that different parameter values are found when focussing on peaks or low flows.

10. Table 2: The b parameter values for the two study catchments seem to be well outside the expected range of values shown in the table. How can this be explained? Do the parameters compensate for modelling errors during the calibration process?
We think this difference can be explained by deviations from laboratory conditions. In the Cabauw polder, only the top 70 cm of the soil was clay. Below the clay, the soil is mostly peat, with different soil physical properties. In the Hupsel Brook catchment, some layers can be found with more organic material, gravel or loam. In both catchments, macropores (cracks and animal burrows) were present.

11. Fig. 1: I am not sure this figure is very useful, but I leave the authors decide whether it should be kept in the paper.
We added Figure 1 to show that the model is not only useful in The Netherlands, which is known as a lowland and sometimes considered exceptional, but in many places around the world. We added to the caption: "Lowland areas can be found all over the world and often in densely populated areas, which shows the relevance of WALRUS for application outside The Netherlands".

12. Fig. 4: Why dv appears twice in the vadose zone?
We wrote d_V twice to show that it can be explained both as the lack of water in the unsaturated zone (the area right of the soil moisture profile) and as an effective depth.

13. Fig. 11: Add the symbols on the lines in the legend, so the differences between lines are clear if the article is printed black and white.
We changed the lay-out of the Figure. The contents of the Figure also changed slightly. We made the Figure with a previous version of the model code, which has small differences with respect to the final model code reported in the manuscript. When we run the model again with the correct model code, the Figure changes slightly. The conclusion remains the same, but to the sentence "Note the erroneous time delay and magnitude of the discharge peak when no substeps are used", we added ", in particular for the three hourly time step'.'

Cited references

Beven, K. and P. Young (2013). A guide to good practice in modeling semantics for authors and referees. Water Resources Research 49(8): 5092-5098.

Reviewer 2

This paper presents a rainfall-runoff model for use in lowland areas, the Wageningen Lowland Runoff Simulator (WALRUS). The study focuses on the development of the model, the model structure and how the model is implemented. The development of novel rainfall-runoff models is an interesting topic of research, particularly for lowland areas. It was clear from the paper that the authors had put a lot of thought into what they wanted to include in their model and how this would be formulated within the model structure. Overall the paper is well written and the figures are well presented. However, there are a couple of issues that need to be addressed, which would help strengthen the message of the paper and highlight the novelty of the model you have developed.

General comments:

1. I feel the paper is very long for a description of a new

rainfall-runoff model and as a result, the interesting parts of your paper tended to get a little lost.

We developed the model using 3 sources of information: (1) the discussion on challenges in rainfall-runoff modelling (with e.g. overparameterization), (2) our experience from the two field sites described in detail in the companion paper and briefly in this paper and (3) a literature review of the challenges of modelling in lowland catchments. We changed the set-up of the first Sections to emphasize these contributions to the developing process. We also explained this at the end of the introduction: "First, we describe the threefold motivation for model development: general challenges in rainfall-runoff modelling (Section 2), the two contrasting lowland field sites which were used in the model development (Section 3) and challenges in modelling rainfall-runoff processes in lowland catchments (Section 4."

Here are my suggestions to improve the readability of the paper :

a. p. 1360 (lines 13 -28) 1361 (lines 1 17) I feel you could significantly shorten this section into a single line. These issues are well known and do not require a detailed explanation.

We agree that these subjects are well-known among hydrologists. However, we expect this paper to be published in a general geoscientific journal and may be read by people from different fields. Therefore we feel that it is necessary to include this short discussion on the major challenges in rainfall-runoff modelling for non-hydrologists who are not familiar with the topic. We did move this discussion to a separate Section (Section 2), making it easy for people who are familiar with the material to skip this discussion.

b. I would shorten section 3. I found some of the literature review on the lowland specific hydrological processes overly long and I also felt there was a lot of repetition between section 3 and section 4 when describing what you had implemented in WALRUS. Although these aspects are important, it would improve readability if they were described more briefly.

We changed the set-up of Section 4. We ended each subsection of Section 3 with a summarizing problem definition and moved the last paragraphs of the subsections (explaining how these problems were solved in WALRUS) to the place where that model component is explained.

2. I felt the novel aspects of your model were not highlighted strongly enough. In your conclusion, it would be useful to end with a very clear statement detailing what is novel about your model, how it is different to the previous model formulation and how this advances hydrological modelling within lowland regions.

We stressed the unique selling points in the conclusion by changing "The model includes..." to "The model explicitly accounts for processes which are typically not included in parametric rainfall runoff models, but which are important in lowland areas...".

We also added the following paragraph at the end of the conclusion: "Compared to other rainfall-runoff models, WALRUS has some important advantages: it (1) is applicable to both freely draining and polder areas, (2) is computationally efficient, (3) has few parameters (only 4 to calibrate), (4) has a clear (qualitative) relation between model states and measurable variables, (5) has default options for initial conditions and parameterizations (which can easily be changed for research purposes), and (6) is open source and freeware (programmed in R). These advantages make WALRUS suitable for operational flood and drought forecasting, real-time control, input for hydraulic models, risk analyses, scenario analyses, infrastructure design and time series gap filling in lowland catchments. In that sense, WALRUS complements existing rainfall-runoff models, containing the core hydrological processes for lowlands, while maintaining a simple model structure. This makes WALRUS suitable for discharge simulations by researchers and practitioners alike."

3. I also took a quick look at the companion paper in HESSD and have a couple of comments:

a. I felt there was a lot of repetition between the two papers in the first paragraph of the introduction. You may want to consider changing this so there is a clear distinction between the two papers.

The two papers are companions, but should also be readable separately. Therefore we decided to include a short summary of the GMD-paper in the HESS-paper to introduce the model and its special features.

b. I agree with Reviewer #1 that a comparison between the Wageningen model and the new WALRUS model would make a useful contribution to the HESSD paper. It is difficult to tell how improved your new model formulation is if there is no comparison with the old one.

WALRUS is not an improved version of the Wageningen Model, but rather a brand new model inspired by the Wageningen Model. An intercomparison study is a good idea, but we would prefer to compare it not only the the Wageningen Model, but also to other conceptual rainfall-runoff models. This would be outside the scope of the HESS paper, but we hope we can perform this research in the near future.

Specific comments:

1. p. 1358 (line 20) Lack of topography doesnt make sense, maybe change it for low-lying topography

We changed this sentence to "low elevations and mild slopes"

2. p. 1362 (lines 6 12) and elsewhere in the text. Change all instances of sect to section so that this is consistent.

Done.

3. p. 1362 (lines 24 26) Over what time period was the groundwater and soil moisture data collected?
We added "period 1976-1984".

4. p. 1363 (line 6) we used data of 1993 should be we used data from 1993
Done.

5. p. 1366 (lines 12 - 14) Sentence is unclear and needs re-writing. We rewrote the sentence.

6. p. 1371 (line 5) depending should be changed to dependent
Done.

7. p. 1372 (lines 6 13) I wasnt entirely convinced by this what is the physical reasoning for 25mm bins? Would this relationship change if you increased/decreased the size of the bins? The authors might want to expand on the implications of condensing this information into a single equation given the wide scatter in the data.
Using all points in the plot yielded nearly the same results. We added the bins to help visualise the relation between evapotranspiration reduction and storage deficit. The scatter is large and the uncertainty around this relation is appreciable, which is a reason for not using a more complex parameterisation with more than two parameters.

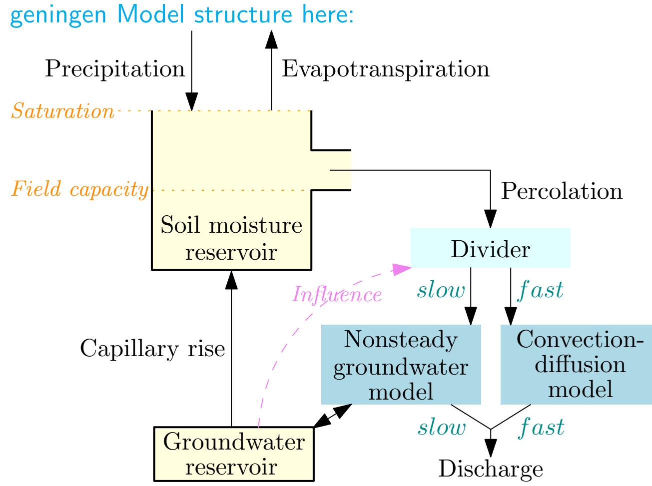
8. p. 1378 Section 4.10. From the HESSD paper, it appears that you have a time series of surface water supply for the Cabauw catchment. Why did you decide to use an artificial event here to illustrate the model formulation rather than showing how it performed for real data? I found the accompanying plot to this experiment (Figure 9) quite confusing and I had to go back to the text a lot to understand what it was showing. It would be worthwhile making the caption and figure legend clearer.
We decided to use artificial events to test if the model produces the feedbacks we intended. In real situations, it is hard to see the effect of a certain model element, because natural variations in the forcing capture the eye. We expanded the captions of the plots to include more information of the variables.

We changed the first part of the caption of Figure 5 to: "The isolated effect of including a wetness-dependent divider between slow and quick flowroutes. Results of a numerical experiment with (solid) and without (dashed) a variable divider (W; the wetness index). A change in W does not only affect quickflow (f_{QS}), but propagates through the model and alters nearly all model variables: storage deficit (d_{V}), groundwater depth (d_{G}) and discharge (Q)."

We changed the caption of Figure 9 to: "Results of a numerical experiment with (solid) and without (dashed) using the surface water level (h_{S}) in the groundwater drainage flux (f_{GS}) computation. Right panels also include the effect of surface water supply (f_{XS}). For the dashed lines in the left panels, h_{S} was computed without f_{XS} and f_{XS} was added to the discharge (Q) afterwards."

9. I believe you could remove Figure 1 it does not add to the paper.
We added Figure 1 to show that the model is not only useful in The Netherlands, which is known as a lowland and sometimes considered exceptional, but in many places around the world.

10. I thought Figure 4 was really well presented. However, I would find it useful to have an overview of the Wageningen model structure alongside the new model formulation to gain a really clear picture of what has changed between the two models. This would help to support what is already written in the text.
We intentionally did not include a figure with the model structure of te Wageningen Model to avoid confusion. WALRUS is only based on the Wageningen Model in the sense that we used experience gained with the Wageningen Model to develop WALRUS. To show the large differences between the two model structures, we include a Figure of the Wageningen Model structure here:



11. Figure 7 and Table 2. Why do the theoretical curves and parameters deviate so much from the actual measurements and fitted parameters (especially b)?
We think this difference can be explained by deviations from laboratory conditions. In the Cabauw polder, only the top 70 cm of the soil was clay. Below the clay, the soil is mostly peat, with different soil physical properties. In the Hupsel Brook catchment, some layers can be found with more organic material, gravel or loam. In both catchments, macropores (cracks and animal burrows) were present.

What would be the implications if you did not have access to this data and instead used theoretical values?
In the HESSD-paper we investigated the effect of using a relation with parameters from the Brooks-Corey table (Figure 14). The effect was large when the parameter set calibrated

with one $d_\mathrm{G}-d_\mathrm{V}$-relation was used with a different $d_\mathrm{G}-d_\mathrm{V}$-relation (top graph), but the differences were minor when the model was calibrated again (bottom graph).