Detailed response to the comments of referee 1

We want to thank referee 1 (M. Herrnegger) for his accurate and helpful review of our manuscript. In this author comment, we list how each of the remarks provided by the referee was addressed. The comments made by the referee will be referred as RC and printed in bold; the authors' comments and answers as AC.

Concerning general comments:

1 RC: The authors show interesting and relevant simulation results that are in good agreement with data sets not used in calibration (FSC, SWE, ETA). These detailed results are however only presented for 6 out of 50 catchments. It is clear that (i) snow is not relevant in all catchment, (ii) availability of SWE measurements may be limited and (iii) that the ETA estimates from MOD16 are unreliable in alpine regions. It would however be of interest to see results of the model performance for more catchments in this context.
AC: We agree. We propose to extend the results to 8 catchments for FSC (a compromise between nival behaviour and data availability). For AET, we also propose to extend the results to 8 catchments (a compromise between pluvial behaviour and data availability). Unfortunately for SWE, we are not able to present other results due to data availability of at site snow gauges.

2 RC: The paper reports on different versions of the hydrological model MORDOR. It would therefore be important for more clarity to add two overview tables, containing (i) model parameters, units, range of parameter values and description and (ii) model fluxes and states, also including units and a description of the variables.
AC: To address this issue we added a supplementary table in Appendix A which summarize MORDOR V1/SD free parameters, units, prior range and description. In addition we completed the description of model fluxes and states in Appendix A. Table 1 was also improved. On the other hand, concerning historical model version (MORDOR V0) we only added explicit references to existing publications which describe the model.

3 RC: MORDOR v1 and SD include a modified and improved snow routine. How is snow sublimation considered in the models?
AC: Snow sublimation is not taken into account in the models. Although it can be a significant process at local scale, it is considered to be of second-order at catchment scale in our regions. See for example Strasser et al. (2008).

4 RC: From the manuscript it is unclear, what temperature data is used
AC: The temperature data used in the study are gridded and provided by (Gottardi et al. 2012). These data result from a statistical reanalysis based on ground network data and weather patterns (Garavaglia et al. 2010). They are available for the 1948-2012 period at 1-km / 1-day resolution. See P3L3-6.

5 RC: How is PET / ET0 estimated? How large are the differences

between the PET values in version v0 compared to version v1/SD? This is crucial since it will influence the AET results.
AC: PET is estimated differently for the three model structures.
For V0, PET is calculated from a statistical formulation driven by air temperature T, as follows:

$$PET = a.(T - b)^2$$

with a and b two free parameters which are calibrated with the other model parameters.
For V1/SD, PET may be estimated by any PET formula. In this study, we use the Oudin formulation, which is expressed as follows:

$$PET = 0.408.Rpot.\frac{T + 5}{100}$$

with $Rpot$ the potential solar radiation ($MJ.m^{-2}.d^{-1}$) and T the air temperature (C).
On the study catchments, Oudin PET vary from about 420 $mm.yr^{-1}$ to 890 $mm.yr^{-1}$.
On the other hand, MORDOR V0 calibrated PET vary from about 220 $mm.yr^{-1}$ to 1750 $mm.yr^{-1}$.
We agree with the reviewer, these differences in the PET estimates obviously have a great impact on AET results. We added a specific comment in section 4.3.

6 RC: What is the reason to use KGE in the objective function and NSE for evaluation?
AC: We use the KGE for calibration because of its good statistical properties, which are helpfull for parameters identification. On the other hand, model evaluation is based on NSE because this criterion is commonly used for evaluation of hydrological models and is therefore suitable to use as a benchmark for this study. In addition, it allows to consider different metrics for calibration and posterior evaluation.

7 RC: The authors state that the model runs at different temporal resolution. Are the calibrated parameters comparable between the different temporal resolutions?
AC: Most parameters do not depend on temporal model resolution. However some parameters like the unit hydrograph parameters and the parameters used in the L storage outflow equation remain dependent on the temporal resolution. Concerning the model calibration process, we use a wide prior range for parameters values which is relevant for both daily and sub-daily time steps.

8 RC: In the Appendix the model formulations are given. Here it would also be interesting to give some technical details on the models: In what language are they written? Is there an internal time discretisation implemented? How long does a run take?
AC: We propose to add this paragraph in the Appendix : "The MORDOR SD model is written in FORTRAN 90. The model runs at different temporal resolution. The duration of a simple model simulation (i.e. model run and evaluation criteria computation) is about 1 sec and depends on the time step and on the length of time series. For instance a daily simulation over 50 years takes less than 1 sec and an hourly simulation over 10 years takes about 2 sec. Concerning

the calibration process (about 40.000 model runs), the algorithm takes about 10 min for a daily time step over 50 years and about 45 min for an hourly time step over 10 years. Post-processing and graphical tools are developed in R language."

Concerning specific comments:

RC: P1L22: semi-distributed
AC: It has been changed in the revised manuscript.

RC: P2L1: Most studies
AC: It has been changed in the revised manuscript.

RC: P2L18-19: It is unclear why the two best solutions are selected (I presume the presented MORDOR v1 and SD) and why these three new formulations are then compared with the historical version.
AC: "Three" has been replaced with "two" in the revised manuscript.

RC: P2L23: . . . in quality and length of available records.
AC: It has been changed in the revised manuscript.

RC: The number of free parameters are not underlined in the table.
AC: We propose a new redaction of the legend of Table 1 : ". . . For each module and model, the number of free parameters is given."

RC: Table 1: Could you please clarify what is to be understood under "adjusted PET from a statistical formulation driven by temperature"?
AC: See response to RC 5.

RC: Figure 1: It would be good to highlight and name the catchments shown in detail in the analysis.
AC: We propose to provide as Supplementary Material : (i) a table of the main features of the catchments used, including name, geographical position (i.e. coordinates of the outlet), catchment area, elevation range, slope, annual P, annual PET, annual Q, time step, modelling periods P1 and P2; (ii) a specific figure with the catchments location.

RC: P5L23: Why not call the section 3.1.2 simply MORDOR V1
AC: We propose to homogenize the three sections titles as follows:
3.1.1. MORDOR V0: Initial lumped formulation
3.1.2. MORDOR V1: Revised lumped formulation
3.1.3. MORDOR SD: Semi-distributed formulation

RC: P5L24: revised model formulation
AC: It has been changed in the revised manuscript.

RC: P6L18: . . . time-step based on air temperature.
AC: It has been changed in the revised manuscript.

RC: Figure 2: A subdivision in 10 elevation zones is shown. Is this the standard number of zones used for spatial discretisation, also for the other catchments?
AC: No. In this study, the number of elevation zones depends on the hypsometric curve of the catchment according to the following criteria: (i) the relative area of each elevation zone has to be greater or equal to 5% and less or equal to 50%, (ii) the elevation range of each zone has to be lower than 350m.
We propose to add this comment in section 3.1.3.

RC: P8L7: The term "streamflow interannual daily regime" is not easily understandable.
AC: We agree with your comment, so we propose a new formulation for the section 3.2.2, as follows:

"The runoff signatures are viewed in such a way that streamflow data may be broken up into several samples, each of them a manifestation of catchment functioning. Five different signatures are used in this study and are described in the following:

- time serie of flow is obviously the first signature which has to be reproduced by the model (hereafter called Q);
- long-term mean daily streamflow is used to focus on the capacity to reproduce seasonal variation of observations (hereafter called $Qsea$);
- flow duration curve focuses on the capacity to reproduce streamflow variance and extremes (hereafter called FDC);
- flow recessions during low flow period focuses on streamflow recessions (hereafter called $Qlow$);
- $lag - 1$ streamflow variation is the last signature focusing on short term variability (hereafter called dQ and computed as follows: $dQ(t) = Q(t) - Q(t-1)$).

RC: P8L8: Do you mean the flow duration curve with "streamflow empirical cumulative distribution"?
AC: Yes. See comment above.

RC: P8L9: How are the streamflow recession periods defined in practice?
AC: Streamflow recessions sequences are extracted from the low flow period for each catchment. We only select recessions with a minimum duration of 7 days.

RC: P8L10: What is meant with the 1 st -lag streamflow derivates and how are they calculated?
AC: Yes. See comment above.

RC: P8L11: ... performance results are resumed using... What is meant with resumed?
AC: We replace "resumed" by "summarized"

RC: P8L27: (i.e. 100 simulations)
AC: It has been changed in the revised manuscript.

RC: P9L8: in figure 4
AC: It has been changed in the revised manuscript.

RC: Figure 4: Legend is missing
AC: We added the legend in figure 4.

RC: Figure 7: Why is R shown and not NSE (as in other plots)?
AC: Your remark is relevant. We therefore change the figure criteria, with NSE instead of R. The conclusion remains the same.

P12L7: (see Figure 2?)
AC: It has been changed in the revised manuscript.

RC: P12L20: What is the second order impact? Not being important?
AC: We replaced by "... has a second order effect".

Concerning Appendix A:

RC: Additional tables as mentioned above showing parameters, fluxes and states would very much improve clarity when reading the appendix. It is very frequently unclear what the used variables mean. Figure 10 does not help very much in this context.
AC: See response to comment RC 2.

RC: P13L31: What is meant with "flow length of each gridcell to the outlet", since we are talking about zones?
AC: The $\bar{fl}$ parameter is the average of the flow length of each DEM pixel to the outlet. This parameter is used in the routing function (diffusive wave equation, see A.6) that remains lumped, i.e. there is no relation between $\bar{fl}$ and the elevation zones.

RC: Eq. A1 & A2: Please check equations. They seem to be erroneous.
AC: We checked and we corrected equation A2.

RC: Eq. A3: Why ET0 and not PET? This is not consistent with the other parts of the manuscript.
AC: We agree. We propose to replace ET0, which has a very precise acceptation, by PET.

RC: Eq. A4: Should Kc and Rpot not have a time component?
AC: We agree. We changed equation A4 and A5.

RC: Eq. A5: It is unclear, for what A5 is needed.
AC: A5 equation is needed to have a mean value of K_C equal to 1, whatever the $kmin$ parameter value. In fact, $kmin$ is a shape parameter representing the

seasonal variability of the vegetation, and we want to reduce the sensitivity of the evapotranspiration amount to $kmin$.

RC: Eq. A11: Kf : Is the time component missing? It is unclear, what the difference between kf and kfp is.
AC: K_f is a time variable coefficient. We therefore modified equations A10 and A11. We also added a specific comment to clarify k_f and kf_p terms.

RC: Eq. A14: From the equations it is unclear, how ui(t) is calculated.
AC: As illustrated on Figure A2, $u_i(t)$ is calculated as follows :

$$u_i(t) = u_i(t-1) + in_{U,i}(t) - evu_i(t)$$

RC: Eq. A21: From the equations it is unclear, how zi(t) is calculated.
AC: As illustrated on Figure A2, $z_i(t)$ is calculated as follows :

$$z_i(t) = z_i(t-1) + in_{z,i}(t) - evz_i(t)$$

RC: Eq. 23/25: What is si ?
AC: s_i is the relative area of the zone i, see P14L1

RC: Eq. 24: What is k_n ?
AC: The parameter k_n is the outflow coefficient of the deep storage N. We added a specific comment in Appendix A.

RC: Eq. 26: What is fl?
AC: fl is the mean of flowlength of each gridcell to the outlet [km], see P13L31 and comment above.

References

Strasser, U., Bernhardt, M., Weber, M., Liston, G. E., Mauser, W. (2008). Is snow sublimation important in the alpine water balance?. The Cryosphere, 2(1), 53-66.

Detailed response to the comments of referee 2

We want to thank M. Hrachowitz for his accurate and helpful review of our manuscript. In this author comment, we list how each of the remarks provided by the referee was addressed. The comments made by the referee will be referred as RC and printed in bold; the authors comments and answers as AC.

1 RC: The manuscript will benefit from being proof-read by a native English speaker to reduce the number of typos and language errors (grammar, syntax and use of specific words/terms).
AC: To answer your suggestion, the final manuscript will be checked by native English speaker.

2 RC: It will be of tremendous help for the reader if the author provided tables of (a) the catchments used (including names, geographical positions, catchment areas, elevation range, slopes, annual P, annual potential E, annual Q, modelling time period, and time step (b) the parameters of each model, the associated symbols, units, prior distributions (are these the same for all catchments?) and descriptions (c) all model components (i.e. states and fluxes), including their symbols, dimensions and descriptions. This would make it much more convenient to follow the Appendix, in which many symbols are not clearly defined at this point. If deemed suitable, these tables can be provided as Supplementary Material.
AC: We agree. We propose to add as Supplement Materials a specific section (S1) that presents more in details the dataset of the 50 catchments. Table S1 presents the main features of the catchments dataset, including name, geographical position, area, elevation range, slope, annual P, annual PET, annual Q, time step, modeling periods P1 and P2. Concerning model description and parameters we added a supplementary table in Appendix A which summarize MORDOR V1/SD free parameters, units, prior range (the same for all the catchments) and description. In addition we completed the description of model fluxes and states in Appendix A. Table 1 was also improved. On the other hand, concerning historical model version (MORDOR V0) we only added explicit references to existing publications which describe the model.

3 RC: Section 3.2.2 will benefit from a clearer description of the different criteria. For example, it remains unclear what is meant by "streamflow regime". I suppose it is the long-term seasonal pattern, but please make this more specific. Similarly, the cumulative distribution of flows is commonly referred to as flow-duration curve. A more consistent terminology will help the reader to better appreciate the manuscript. It is also not clear what is meant by 1st-lag flow derivative. Does this refer to the lag-1 autocorrelation? Of flows? Of the recession? Please elaborate!
AC: We agree. We reformulate section 3.2.2 as follows :
"The runoff signatures are viewed in such a way that streamflow data may be broken up into several samples, each of them a manifestation of catchment functioning (Euser et al., 2013; Hrachowitz et al., 2014; Westerberg and McMillan, 2015). Five different signatures are used in this study and are described in the

following:

- time serie of flow is obviously the first signature which has to be reproduced by the model (hereafter called Q);
- long-term mean daily streamflow is used to focus on the capacity to reproduce seasonal variation of observations (hereafter called $Qsea$);
- flow duration curve focuses on the capacity to reproduce streamflow variance and extremes (hereafter called FDC);
- flow recessions during low flow period focuses on streamflow recessions (hereafter called $Qlow$);
- $lag-1$ streamflow variation is the last signature focusing on short term variability (hereafter called dQ and computed as follows: $dQ(t) = Q(t) - Q(t-1)$).

To go further, model realism is also evaluated in regards to three other hydrological variables: (i) fractional snow cover (FSC); (ii) snow water equivalent (SWE); (iii) actual evapotranspiration (ET).

However, observations available for these variables suffer from many limitations and uncertainties (see section 2). Consequently, a specific evaluation is conducted and is explained in sections 4.2 and 4.3."

4 RC: The post-calibration evaluation of the models with respect to snow and evaporation dynamics is an important point in this paper. Yet, no mention of this is made in section 3.2.2. How are MODIS data used to compare to model output? Spatial averages? What about the temporal resolution of the evaluation? Which performance metric was used? Some of this is mentioned later in the manuscript but I think this needs to be made clear in the methods section.
AC: We agree. We propose to add a specific comment in section 3.2.2 to mention the other hydrological variables, see response to comment 3 above. In addition we added more details in sections 4.2 and 4.3 to clarify how the satellite data (MOD10 and MOD16) are used.

5 RC: Related to (4), I did not understand how a fractional snow cover can be reproduced with lumped model formulations (VO and V1). This makes clearly sense for a semi-distributed model (Mordor SD). But obviously I missed something for the lumped versions. Please clarify!
AC: For the lumped versions of the model, the fractional snow cover FSC estimates are based on a statistical formulation founded on the hypsometric curve. More in details, during accumulation the FSC is computed according to a monotonic crescent function as follows:

$$FSC(t) = 1 - \frac{\arctan \frac{\gamma(t) - fp1}{fp2} + fp3}{fp4} \tag{1}$$

where the parameters $fp1$, $fp2$, $fp3$ and $fp4$ vary from a catchment to an other as a function of the hypsometric curve and the orografic gradient. The state variable $\gamma(t)$ depends on the snow pack $S(t-1)$, its temperature $t_S(t-1)$, the snow $N(t)$ and its temperature $t_N(t)$ as follows :

$$\gamma(t) = \frac{S(t-1) \cdot t_S(t-1) + N(t) \cdot t_N(t)}{S(t-1) + N(t)} \tag{2}$$

For melt, $FSC(t)$ depends on the previous FSC and the evolution of the snow pack as follows :

$$FSC(t) = FSC(t-1) \cdot (\frac{S(t)}{S(t-1)})^{0.5} \tag{3}$$

6 RC: What is the reason behind using KGE for calibration (which is completely fine) but NSE for evaluation? Why is not the same metric used for both?

AC: We use the KGE for calibration because of its good statistical properties, which are helpfull for parameters identification. On the other hand, model evaluation is based on NSE because this criterion is commonly used for evaluation of hydrological models and is therefore suitable to use as a benchmark for this study. In addition, it allows to consider different metrics for calibration and posterior evaluation.

7 RC: The presentation of the results and discussion section would strongly benefit from a bit more detail. Detailed results are only shown for a few catchments with good overall performance. And even for these, it remains unclear how the modelled hydrograph looks like (in comparison to the observed one) and what the values of the individual associated calibration objective functions (i.e. the 3 individual KGEs) and evaluation metrics(the remaining criteria) are. In addition,I think it would also be valuable to show examples of catchments where the model adaptation did not work and also discuss why.

AC: We propose to add as Supplement Materials two specific sections about the model comparison over calibration periods (S2) and over evaluation periods (S3). The section S2 presents more in details the performance of MORDOR V0, V1 and SD over the calibration periods P1 and P2. For the three considered models, we show values of the individual associated calibration metrics (KGE(Q), KGE(Qsea) and KGE(FDC)) for all the catchments over the calibration periods P1 and P2. In addition we show, for each of the 50 catchments, the observed hydrographs and those modeled by MORDOR V0, V1 and SD (calibration mode). Similarly, section S3 presents more in details the performances of MORDOR V0, V1 and SD over the evaluation periods P1 and P2. For the three considered models, we show the values of the individual associated evaluation metrics (NSE(Q), NSE(Qsea), NSE(dQ), NSE(FDC) and NSE(Qlow)) for all the catchments over the evaluation periods P1 and P2. We show also, for each of the 50 catchments, the observed hydrographs and those modeled by MORDOR V0, V1 and SD (evaluation mode). Concerning snow and evapotranspiration processes, we extend the analysis to other catchments, with 8 nival catchments for FSC and 8 pluvial catchments for AET.

8 RC: Related to(7), it is mentioned that V1 provides substantial improvements compared to V0. As V1 is changed in various respects

in comparison to V0, it would be great if the authors invested a bit of effort to analyze and document which part/adjustment of V1 contributes most to the improvement.
Ac: Mordor V1 differs from V0 especially for water balance formulation and snow modelling. However is very difficult to trace the origin of the various improvements. Logically, for nival catchments the changes in snow modelling (and also the semi-distribution, see figure 4) are efficient. Concerning the water balance and so the improvement in the representation of evapotranspiration processes, we propose to add a specific comment in section 4.3 in order to analyze the origin of AET differences.

9 RC: P.1,l.6: what is meant by "inflected"? Please rephrase.
AC: Is has been changed in the revised manuscript.

10 RC: P.1,l.8: should read as "...evapotranspiration estimates. The model comparison is...."
AC: It has been changed in the revised manuscript.

11 RC: P.1,l.22: should read as "...semi-distributed..."
AC: It has been changed in the revised manuscript.

12 RC: P.1,l.23: Nijzink et al. 2016 would fit in nicely here.
AC: Good suggestion. We added the reference in the text.

13 RC: P.1,l.23: what is meant by "To overpass hydrological singularity...."? Please rephrase.
AC: It has been changed in the revised manuscript.

14 RC: P.2,l.8: I may be worth referring to Hrachowitz et al. (2014) here.
AC: Good suggestion. This paper is well suited to this context and we added the reference in the text both in introduction and in section 3.2.2.

15 RC: P.2,l.15: should read as "...framework on the MORDOR...."
AC: It has been changed in the revised manuscript.

16 RC: P.2,l.20-22: irrelevant. Can be condensed.
AC: We agree. We removed this paragraph.

17 RC: P.2,l.26: should read as "......mainly in the Alps (18 catchments), the Pyrenees (5 catchments) and the Massif Central..."
AC: It has been changed in the revised manuscript.

18 RC: P.2,l.29: should read as "...hydrological conditions. The average area of the study catchments is..."
AC: It has been changed in the revised manuscript.

19 RC: P.3,Table 1: not clear if the 22/17/19 parameters are all calibration parameters, as it seems in the Appendix that some of them are fixed. Please clarify.

AC: We tried to clarify this in Table 1. See comment to RC 2.

20 RC: P.3,l.4: should read as "...1635 mm/yr. With regard to...."
AC: It has been changed in the revised manuscript.

21 RC: P.3,l.10: should read as "...sub-daily time steps..."
AC: It has been changed in the revised manuscript.

22 RC: P.3,l.11: what is meant by " the shape of local gauges"? Please clarify.
AC: We agree. We propose a new formulation : "the hourly records of locals gauges are used to compute areal precipitation and temperature at 12-, 8- and 6-hours time step."

23 RC: P.3,l.12: that is ok, but it should be underlined that these are not observations but modelled estimates which can be subject to considerable uncertainty.
AC: We agree. We propose a new formulation: "It has be noticed that these data are not observations but modelled estimates which can be subject to considerable uncertainty."

24 RC: P.4,l.1-2: should read as "...for being affected by many..."
AC: It has been changed in the revised manuscript.

25 RC: P.4,l.5: should read as "...provides fractional snow cover..."
AC: It has been changed in the revised manuscript.

26 RC: P.4,l.5: please explain what "fractional snow cover" describes. Are these spatial fractions? If yes across the entire catchment? Across a pixel? Which value was used to compare the modelled values with?
AC: The satellite MOD10 product provides gridded snow cover time-series. In this study we average the gridded values at catchment scale in order to compute a fractional snow cover. It has been rephrase in the revised manuscript.

27 RC: P.4,l.15: should read as "...interconnected storages."
AC: It has been changed in the revised manuscript.

28 RC: P.4,l.15: what is meant by "continuously"? Please clarify.
AC: With this phrase we want to emphasize that MORDOR is a continuous hydrological model and not a event-based model. We propose a new formulation: "Is is a continuous model that can be can be used with a time step ranging from hourly to daily."

29 RC: P.5,l.1: No, what is required is a *representative* estimate of areal precipitation. The mean (or any other measure of central tendency) will average out extremes, which will,due to the non-linear nature of your(or better: any meaningful hydrological model), result in biased results.
AC: We agree, the term "mean" is improper in this context. It has been changed

in the revised manuscript.

30 RC: P.5,l.19: "(ii) snow modelling have to be improved..." reads awkward. Please rephrase.
AC: We rephrased as follows: "...representation of snow processes have to be improved..."

31 RC: P.5,l.31: should read as "...evapotranspiration, the model..."
AC: It has been changed in the revised manuscript.

32 RC: P.5,l.32: what is meant by "neutralized"??
AC: We propose a new formulation as follows: "(i) a surface interception: net rainfall and evapotranspiration capacity are calculated from the subtraction of MET from rainfall"

33 RC: P.6,l.11: It is not clear which part of the system the ground-melt component represents. What exactly does it do? Please clarify.
AC: The snow ground melt corresponds to the melting component coming from the ground heat flux, see for example (DeWalle, D. R., Rango, A.,2008). From experimental studies (e.g. Whitaker and Sugiyama, 2005), the ground-melt rates range form 0.5 to 1 mm/day.

34 RC: P.6,l.30: that is fine, but please specify if the gradients are set to fixed values or if they are calibrated (similar to rainfall multipliers). Where do the values (fixed or prior distributions) come from? Literature? Please provide references.
AC: The orographic gradients are calibrated with a uniform prior distribution whose upper and lower limits come from the climatic reanalysis used in this study (Gottardi et al., 2012). See the revised version of table 2 and appendix A.

35 RC: P.7,l.2,section 3.2: I would suggest to rearrange this section for a better flow and to start with the calibration approach, followed by the split sample test and the post evaluation criteria.
AC: We agree. Section 3.2 has been rearranged considering your suggestion. See response to points RC 3 and RC 4.

36 RC: P.7,l.5: does this mean that you end up with 2 parameter sets for each catchments? Is the following analysis then based on these 100 parameter sets (i.e. 2 for each catchment)? Please describe in more detail what you are doing.
AC: Yes we do that as explained in the text, see section 3.2.1. For each catchment we have 2 sets of parameters (θ_1 from P1 and θ_2 from P2) and all the results and performances are calculated from the 100 (2*50) simulations.

37 RC: P.8,l.1-2: this resembles an approach described by Gharari et al. (2013). It would be good to refer to that paper.
AC: Good suggestion. We added the reference in the text.

38 RC: P.8,l.4ff: please clearly separate between criteria that are

used for calibration (i.e. q, reg and qlc) and those used for post-calibration evaluation (i.e. etg, dq, snow cover, evaporation).
AC: We agree. Section 3.2 has been rearranged considering your suggestion. See response to RC 3 and RC 4.

39 RC: P.8,l.17, eq.1: should this not read "KGEqcl"?
AC: It has been changed in the revised manuscript.

40 RC: P.8,l.20: "Numerous applications if this OF..." please provide references.
AC: With numerous applications we refer to industrial studies made at EDF in an operational context. In spite of this, this OF is inspired from Paquet et al. (2013). We rephrased the paragraph in section 3.2.3.

41 RC: P.8,l.29: Do the V1 and SD models in *all* catchments outperform V0 or is it just on average? Please provide some representative examples for both cases of improvements and cases where V1 and SD did not result in improvements.
AC: According to section S3 of Supplement Materials (table S4 and S5), we can observe that for 73% of cases Mordor SD performs better than V0 in regard to streamflow signature. The Agout at La Raviege catchment (period P1) is a good example of important improvement of SD model. On the other hand for Ubaye at RocheRousse catchment Mordor V0 (period P1) performs better than SD.

42 RC:P.9,l.4: the improvement is obvious, but I struggle to see the "spectacular" improvement. In addition,"most" seems also a bit exaggerated here: reg,qcl and etg show only minor improvements, if any. Please tone the statement down a bit to actually reflect what we can see in the figures.
AC: We agree. We qualified our statement as follows : "As a conclusion, the new formulation (V1) provides a significant improvement of performances, specially for q and dq signatures."

43 RC: P.10, Figure 5: are the NSE values the NSE values of the snow cover? Please clarify. In addition, please make sure that *all* figure captions in the manuscript are stand-alone, i.e. that the reader can fully understand a figure only by reading its caption.
AC: Yes. Captions of figures 5, 6,7 and 8 were completed in order to take into account your suggestion.

44 RC: P.10,l.1: what is meant by "overpasses"? please rephrase.
AC: We propose to change "overpasses" by "outclasses".

45 RC: P.10,l.2: what is meant by "...the interest of the..."? please rephrase.
AC: We propose to change "...the interest of the..." by "Therefore, the semi-distributed scheme clearly shows its added value for nival catchments.

46 RC: P.11,figures 6,7: see (43)

AC: See comments above.

47 RC: P.12,l.2: should read as "...that cannot be..."
AC: It has been changed in the revised manuscript.

References

DeWalle, D. R., Rango, A. (2008). Principles of snow hydrology. Cambridge University Press.

Whitaker, A. C., Sugiyama, H. (2005). Seasonal snowpack dynamics and runoff in a cool temperate forest: lysimeter experiment in Niigata, Japan. Hydrological Processes, 19(20), 4179-4200.

Impact of model structure on flow simulation and hydrological realism: from lumped to semi-distributed approach.

Garavaglia Federico[1], Le Lay Matthieu[1], Gottardi Fréderic[1], Garçon Rémy[1], Gailhard Joël[1], Paquet Emmanuel[1], and Mathevet Thibault[1]

[1]EDF-DTG, 21 avenue de l'Europe, BP 41, 38040 Grenoble cedex 09

Correspondence to: Garavaglia Federico (federico.garavaglia@edf.fr)

Abstract. Model intercomparison experiments are widely used to investigate and improve hydrological model performance. However, a study based only on runoff simulation is not sufficient to discriminate different model structures. Hence, there is a need to improve hydrological models for specific streamflow signatures (e.g., low and high flow) and multi-variable predictions (e.g., soil moisture, snow and groundwater). This study assesses the impact of model structure on flow simulation and hydrological realism using three versions of a hydrological model called MORDOR: the historical lumped structure and a revisited formulation available in both lumped and semi-distributed structures. In particular, the main goal of this paper is to investigate the relative impact of model equations and spatial discretization on flow simulation, snowpack representation and evapotranspiration estimation. Comparison of the models is based on an extensive dataset composed of 50 catchments located in French mountainous regions. The evaluation framework is founded on a multi-criterion split-sample strategy. All models were calibrated using an automatic optimization method based on an efficient genetic algorithm. The evaluation framework is enriched by the assessment of snow and evapotranspiration modeling against in-situ and satellite data. The results showed that the new model formulations perform significantly better than the initial one in terms of the various streamflow signatures, snow and evapotranspiration predictions. The semi-distributed approach provides better calibration-validation performance for the snow cover area, snow water equivalent and runoff simulation especially, for nival catchments.

1 Introduction

Hydrological models are widely applied in water engineering for design and scenario impact investigations. Depending on the type of application, the catchment characteristics and data availability, different model conceptualizations and parameterizations are considered. In many cases, the choice of the model is the result of the modeler's experience. However, hydrologists have developed objective and rigorous frameworks to evaluate and improve hydrological models.

A common approach to discriminate different model structures is to conduct model intercomparison experiments. Such experiments have been helpful to explore model simulation performance of lumped (e.g., Duan et al., 2006; Breuer et al., 2009), semi-distributed (e.g., Duan et al., 2006; Holländer et al., 2009) and distributed (e.g., Henderson-Sellers et al., 1993; Reed et al., 2004; Holländer et al., 2009; Smith et al., 2012; Nijzink and Savenije, 2016) models in a consistent way using the same input data. To go beyond specific analyses and provide general conclusions, multi-catchment experiments have been proposed by several authors (e.g. Perrin et al., 2001; Gupta et al., 2014) and are now used extensively. Most studies focus only on runoff modeling performance, since runoff is the main data available at the catchment scale. However, as the runoff data are used for both training the model and its validation, one may question the quality of the prognostic variables produced by the model that have not been optimized through calibration, such as snow, evapotranspiration and soil groundwater (Hrachowitz et al., 2014). Moreover, when focusing only on runoff simulation, we often fail to discriminate different model structures. However, interesting conclusions may be drawn when focusing on particular aspects of streamflow not used in the calibration process: low flows (Staudinger et al., 2011) or high flows (Vansteenkiste et al., 2014) or on other hydrological variables, such as soil moisture (Orth et al., 2015), snow (Parajka and Blöschl, 2006) and groundwater (Motovilov et al., 1999; Beldring, 2002).

In a similar way, this paper compares different model structures in terms of both runoff simulation and hydrological realism. More specifically, we investigate the relative importance of model equations and spatial discretization on flow simulation, snowpack representation and evapotranspiration estimate. This correspondence between model and "reality," often described as "working for the right reasons" (Kirchner, 2006; Kavetski and Fenicia, 2011; Euser et al., 2013), is essential if the model is to be used as a tool for improving the understanding of a hydrological system and/or used for prediction and extrapolation, such as simulating the impacts of land use change, variability in climatological forcing, etc.

We apply this framework oto the MORDOR hydrological model (Garçon, 1996), which has been extensively used by Électricité de France (EDF, the French electric utility company) for more than 25 years for operational applications. Recent changes in the model structure have been made to improve model performance. Many alternative model structures have been tested, which concern both model equations and model spatial discretization, and we selected the two best solutions. In this study we present and compare these two new formulations with the historical version.

2 Data and study area

The comparison of the three hydrological models is based on an extensive dataset composed of data from 50 catchments. This dataset collects different operational case studies from EDF activities. These catchments are located in mountain regions, manly in the Alps (10 catchments), the Pyrenees (5 catchments) and the Massif Central (29 catchments). Four catchments are located in the northeast of France (Ardennes and Jura and Vosges regions), one in the northwest (Brittany region) and one in Corsica. Figure 1 shows the catchment locations. Catchments were chosen based in quality and length of records criteria. The large hydroclimatic range of the dataset ensures the models' consistency in different hydrological conditions.The average area of the study catchments is 911 km², ranging from 20 to 7366 km² and the average of median elevation of the whole dataset is 981 m a.s.l., ranging from 109 to 2365 m a.s.l.

For each catchment the following data were collected: (i) discharge, (ii) rainfall, (iii) temperature, (iv) potential and actual evapotranspiration, (v) fractional snow cover and local snow water equivalent.

The discharge data are provided by EDF and French water management agencies. The average length of records at all these stations combined is around 25 years, ranging from 9 years for Ouveze at Bedarrides (southern Alps) to 53 years for Sioule at Fades (Massif Central). The whole discharge dataset consists of 1526 hydrologic years. The average runoff for the whole

Table 1. Main components of MORDOR V0, V1 and SD models in terms of water balance, runoff production, snow model, routing scheme and spatialization. For each module and model, the number of free parameters is given.

Module	MORDOR V0	MORDOR V1	MORDOR SD
Water balance	Calibrated PET from a statistical formulation driven by temperature. 2 free parameters (see Garçon (1996))	Forced by PET. Crop coefficient formulation. 2 free parameters (see Appendix A3).	
Runoff production	4 storage (U, L, Z, N) and 3 fluxes components (surface, subsurface and base flows). Linear inflow and outflow of storage. 7 free parameters (see Paquet et al. (2013)).	4 storage (U, L, Z, N) and 3 fluxes components (surface, subsurface and base flows). Nonlinear inflow and outflow of storage. 7 free parameters (see Appendix A5).	
Snow model	Snow accumulation driven by the air temperature and hypsometric curve. Classical degree-days formulation for snow melt. 11 free parameters (see Valéry et al. (2014b)).	Snow accumulation driven by air temperature and parametric S-shaped curve. For snow melt: classical degree-days, cold content, liquid water content, ground-melt component and variable melting coefficient. 6 free parameters (see Appendix A4).	
Routing scheme	UH modeled by Weibull distribution. 2 free parameters (see Paquet et al. (2013)).	UH modeled by diffusive wave. 2 free parameters (see Appendix A6).	
Spatialization	None	None	Orographic gradients. 2 free parameters (see Appendix A2).
Total	22 free parameters	17 free parameters	19 free parameters

dataset is around 800 mm/year, ranging from 225 to 1635 mm/year. With regard to forcing data, rainfall and temperature are gridded and provided by Gottardi et al. (2012). These data result from a statistical reanalysis based on ground network data and weather patterns (Garavaglia et al., 2010). They are available for the 1948-2012 period at 1-km² / 1-day resolution. Concerning the rainfall, the average amount for the whole dataset is around 1345 mm/year, ranging from 825 mm/year to 2000 mm/year. The model time step differs from catchment to catchment and depends on hydrological characteristics (area, topography, time to peak, etc.). We modeled 44 catchments at the daily time step, one at the 12-h time step, two at the 8-h time step and three at the 6-h time step. To obtain forcing at the subdaily time step, the gridded data are downscaled according to ground network data at a finer time step, i.e., the hourly records of local gauges are used to compute areal precipitation and temperature at the 12-, 8- and 6-h time steps.

Evapotranspiration data used for validation come from the MOD16 satellite global evapotranspiration product (Mu et al., 2011), which has provided 1-km²/8-day land surface ET datasets since 2000 using the Penman-Monteith equation and a surface

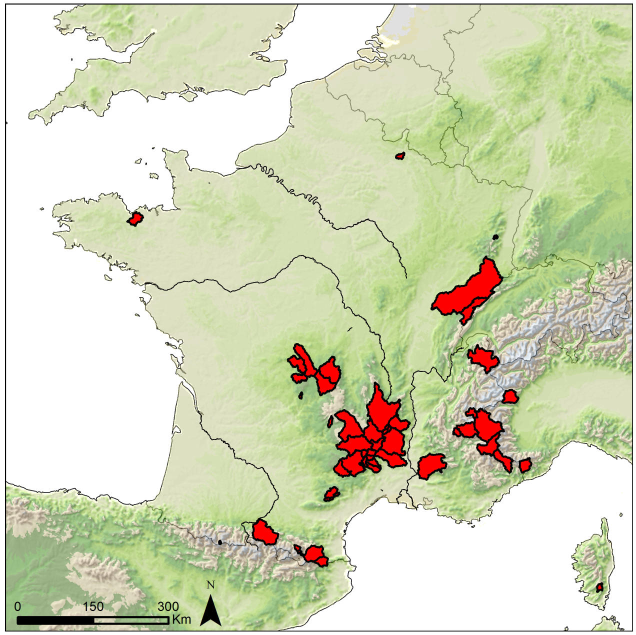

Figure 1. Localization of the catchments studied

resistance derived from MODIS surface data. It has be noticed that these data are not observations but modelled estimates which can be subject to considerable uncertainty. Compared to local flux measurements, the MOD16 product has the great advantage of providing spatially explicit large-scale ET estimates. Some studies have shown its consistency, even if it is known for being affected by many uncertainties, especially in mountainous areas where the global meteorological input is clearly deficient or in tropical and subtropical regions where it clearly underestimates ET (e.g., Trambauer et al., 2014; Hu et al., 2015; Miralles et al., 2015). Two types of snow data are used for model validation: fractional snow cover (FSC) and snow water equivalent (SWE). The MOD10 satellite product (Hall et al., 2002) provides gridded snow cover time-series. This product has been available since 2000 at a 500-m / 1-day resolution, and is widely used for hydrological applications (e.g., Rodell and Houser, 2004; Parajka and Blöschl, 2006; Thirel et al., 2013). In this study we average the gridded values at catchment scale in order to compute a fractional snow cover. Snow Water Equivalent (SWE) data come from the EDF snow network, composed of cosmic ray snow sensors (NRC) (Kodama et al., 1979; Paquet and Laval, 2006). In this study we use three measurement gauges situated within the Durance at La Clapiere catchment: Izoard (2280 m a.s.l), Chardonnet (2455 m a.s.l.) and Marrous (2730 m a.s.l.). SWE time-series at these locations are available since 2001.

3 Methods

3.1 Hydrological model versions

3.1.1 MORDOR V0: Initial lumped formulation

The historical MORDOR model is a lumped conceptual rainfall–runoff model. Its structure is similar to that of many conceptual type models, with different interconnected storage. Is is a continuous model that can be used with a time step ranging from hourly to daily. The required input data are a representative estimate of areal precipitation and air temperature.

The main components of the model are: (i) an evaporation function that determines the potential evaporation as a function of the air temperature; (ii) a rainfall excess/soil moisture accounting storage U that contributes to the actual evaporation and to the direct runoff; (iii) an evaporating storage Z, filled by a part of the indirect runoff component that contributes to the actual evaporation; (iv) an intermediate storage L that determines the partitioning between a direct runoff, an indirect runoff and the percolation to a deep storage N; (v) a deep storage N that determines a baseflow component; (vi) a snow accumulation function calculated from the temperature and the hypsometric curve of the catchment and a rain-snow transition curve; (vii) a snow melt function based on an improved degree-day formulation; and (viii) a unit hydrograph that determines the routing of the total runoff.

In this configuration, the MORDOR V0 model has 22 free parameters (see Table 1) to be optimized during the calibration process. The model was developed in the early 1990s (Garçon, 1996). Since then, it has been extensively used at EDF for operational inflow and long-term water resource forecasting, hydrological analysis and extreme flood estimation (Paquet et al., 2013). Several hundred models have been calibrated in France and abroad (Mathevet and Garçon, 2010). A model inter-comparison study (Mathevet, 2005; Chahinian et al., 2006) based on the assessment of 20 rainfall–runoff models, tested on a sample of 313 catchments at the daily and hourly time steps, has shown that the MORDOR model is among the more efficient and robust rainfall–runoff model structures. Valéry et al. (2014a) also showed that the MORDOR snow module was among the most efficient when compared to six well-known snow modules.

However, various reasons to improve the model have appeared recently: (i) an increase in model performance in terms of floods and low-flow simulations may broaden model applications; (ii) representation of snow processes must be improved to allow for snow data assimilation, particularly for long-term snow melt forecasts; (iii) representation of orographic meteorological variability should be taken into account; and (iv) simplification of the model's structure and parameterization may improve the efficiency of model calibration and reduce parameter equifinality (Beven and Freer, 2001).

3.1.2 MORDOR V1: Revised lumped formulation

The revised model formulation, hereafter called MORDOR V1, does not modify the overall catchment conceptualization. In the following parts, we distinguish changes in: (i) the water balance formulation; (ii) the runoff production; (iii) the snow model; and (iv) the routing scheme. Special focus on the MORDOR V1 components and fluxes is given in Appendix A. In this configuration, the MORDOR V1 model has 17 free parameters to be optimized during the calibration process (see Table 1 and Table 2).

Water balance

The water balance formulation includes a simplified vegetation component, with a maximum evaporation that is derived from the potential evapotranspiration PET using a crop coefficient (Allen et al., 1998). From the maximum evapotranspiration MET, the model calculates actual evapotranspiration (AET) from three components: (i) a surface interception: net rainfall and evapotranspiration capacity are calculated from the subtraction of MET from rainfall (e.g., Perrin et al., 2003); (ii) an evapotranspiration from the root soil, calculated as a linear function of the saturation level of the soil moisture accounting storage U; and (iii) an evapotranspiration from the capillarity water storage in the hillslope, calculated as a linear function of the saturation level of the capillarity storage Z.

Runoff production

The model identifies three flux components: (i) surface runoff; (ii) subsurface exfiltration; and (iii) base flow. Surface runoff is generated by excess water coming from U and L storage. It represents, in a pure conceptual way, both Hortonian and Hewlettian runoff. Subsurface runoff is generated by L storage outflow, calculated as a nonlinear function of the relative saturation. Base flow is generated by N storage outflow, calculated as a nonlinear function of the water content.

Snow and glacier model

The snow model is derived from a classical degree-day scheme, with a few important additional processes: (i) a cold content able to dynamically control the melting phase; (ii) a liquid water content in the snowpack; (iii) a ground-melt component; and (iv) a variable melting coefficient, depending on the potential radiation assumed to model the changing albedo effect throughout the melting season. The accumulation phase is controlled by the discrimination of the liquid and solid fractions of the precipitations. From the temperature, these fractions are derived from a classical parametric s-shaped curve (e.g., Zanotti et al., 2004; Micovic and Quick, 1999). The snowpack is represented by three state variables: (i) the snow water equivalent; (ii) the snowpack bulk temperature; and (iii) the liquid water content in the snowpack. Snow melt is calculated as the sum of superficial and ground melts. Superficial melt is derived from a degree-day formulation, where the melting temperature is snowpack bulk temperature, updated at each time step based on air temperature. A glacier component may also be activated, which is based on a simple degree-day formulation.

Routing scheme

The transfer function is applied to the sum of the runoff contributions. Its formulation is based on the diffusive wave equation (Hayami, 1951).

3.1.3 MORDOR SD: Semi-distributed formulation

The Semi-Distributed MORDOR model is an improvement of the MORDOR V1 model, which includes a spatial discretization scheme. This discretization is based on an elevation zone approach, which is known to be both parsimonious and efficient for mountainous hydrology (Bergstroem, 1975; DHI, 2009). A special focus on MORDOR SD components and fluxes is given in Appendix A. Figure 2 illustrates this discretization on the Durance at La Clapière catchment (2175 km², southern Alps), with 10 elevation zones each representing between 5% and 18% of the total area. In this study, the number of elevation zones depends on the hypsometric curve of the catchment according to the following criteria: (i) the relative area of each elevation zone has to be greater or equal to 5% and less or equal to 50%; and (ii) the elevation range of each zone has to be lower than 350 m.

In most MORDOR SD applications, spatial variability of meteorological forcings is summarized with two orographic gradients, one for precipitation and one for temperature. In this way, we assume that in mountainous areas, spatial variability is mainly driven by altitude. Most of the model state variables are calculated for each elevation zone. Only groundwater water content and outflow are considered as global and are calculated at the catchment scale. In the configuration used in this study, MORDOR SD has 19 free parameters (i.e., 17+2 with two orographic gradients) to optimize during the calibration process (see Table 1 and Table 2).

3.2 Evaluation strategy

3.2.1 Hydrological signatures

The runoff signatures are viewed in such a way that streamflow data can be broken up into several samples, each of them a manifestation of catchment functioning (Euser et al., 2013; Hrachowitz et al., 2014; Westerberg and McMillan, 2015). Five different signatures are used in this study and described in the following:

– the time serie of flow is obviously the first signature that has to be reproduced by the model (hereafter called Q);

– the long-term mean daily streamflow is used to focus on the capacity to reproduce seasonal variation of observations (hereafter called $Qsea$);

– the flow duration curve focuses on the capacity to reproduce streamflow variance and extremes (hereafter called FDC);

– the flow recessions during low flow period focuses on streamflow recessions (hereafter called $Qlow$);

– the $lag-1$ streamflow variation is the last signature focusing on short-term variability (hereafter called dQ and computed as follows: $dQ(t) = Q(t) - Q(t-1)$).

To go further, model realism is also evaluated in regards to three other hydrological variables: (i) the fractional snow cover (FSC); (ii) the snow water equivalent (SWE); and (iii) the actual evapotranspiration (AET). However, these data suffer from many limitations and uncertainties (see section 2). Consequently, a specific evaluation is conducted and explained in sections 4.2 and 4.3.

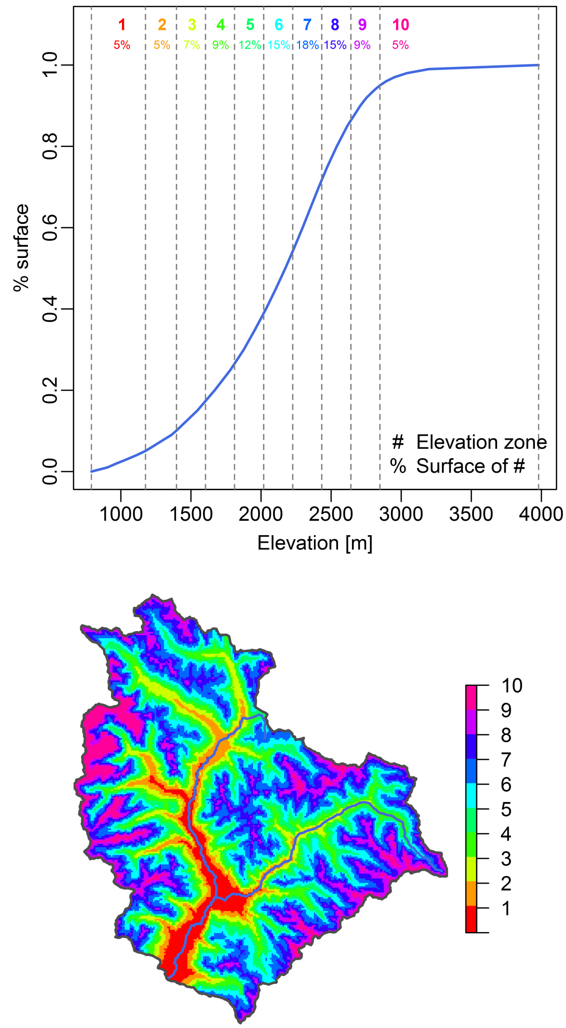

Figure 2. Durance at La Clapiere catchment: (a) hypsometric curve; (b) elevation zones.

3.2.2 Model calibration

The model is calibrated using an efficient genetic algorithm inspired by Wang (1991). This stochastic population-based search algorithm performs approximately 40,000 model runs during a classical calibration process.

The multi-criterion composite objective function (OF) to be minimized during calibration is expressed as follows:

$$OF = (1 - KGE_Q) + (1 - KGE_{Qsea}) + (1 - KGE_{FDC}) \quad (1)$$

where KGE is Kling-Gupta Efficiency (Gupta et al., 2009), which combines three components: correlation, variance bias and mean bias. The triple focus on time-serie, seasonal streamflow and flow duration curve can properly identify the different components of the model. Numerous industrial applications of this OF, within a wide range of hydroclimatic conditions, showed that it was well designed to calibrate the MORDOR model (e.g., Paquet et al., 2013).

3.2.3 Split sample test

To evaluate the model, we adopted the split sample test advised by Klemeš (1986) and Gharari et al. (2013). For each catchment, the entire data record was split into two periods (P1 and P2). In the tests, we first calibrated the models on period P1 and tested them in validation mode on period P2. Then the role of the periods was reversed (calibration on P2 and validation on P1). Therefore, a total of 100 calibrations (50 for P1 and 50 for P2) and 100 validation tests were run on the whole catchment set, and the results were analyzed on this basis.

3.2.4 Evaluation metrics

Model performance is quantified using the classical Nash-Sutcliffe Efficiency (NSE). This criterion is commonly used for evaluation of hydrological models and is therefore suitable to use as a benchmark for this study. In addition, it allows to consider different metrics for calibration and posterior evaluation. NSE criteria are systematically calculated for all the streamflow signatures Q, $Qsea$, FDC, $Qlow$ and qQ and for all the catchments. NSE criteria are also calculated for supplementary hydrological variables (FSC, SWE and AET) but they are not systematically shown, considering the data limitation already mentioned.

4 Results and discussion

This section presents the results of the model comparison. We focus on improvements in terms of model performance and the representation of snow and evapotranspiration processes.

4.1 Improvement of model performance

Figure 3 summarizes the model performance of the three model versions over the validation periods. Distributions of NSE values over the 50 catchments (i.e., 100 simulations) are plotted for the five samples of observations described above (Q, $Qsea$, $Qlow$, $Qlow$ and dQ). It can first be noted that the three model formulations have good overall performance. The $NSE(q)$ values are above 0.8 in validation on more than 80% of the catchments. However, MORDOR V1 and MORDOR SD perform significantly better than MORDOR V0. This is particularly true for Q and dQ signatures. This is less significant for $Qsea$ and $Qlow$ signatures and insignificant for $Qlow$. When considering $NSE(q)$ values, MORDOR V1 and SD have scores above 0.9 for about 10% of the catchments on validation periods. Another interesting result is the very close performance of V1 and SD versions. In conclusion, the new formulation (V1) provides a spectacular improvement in performance on most

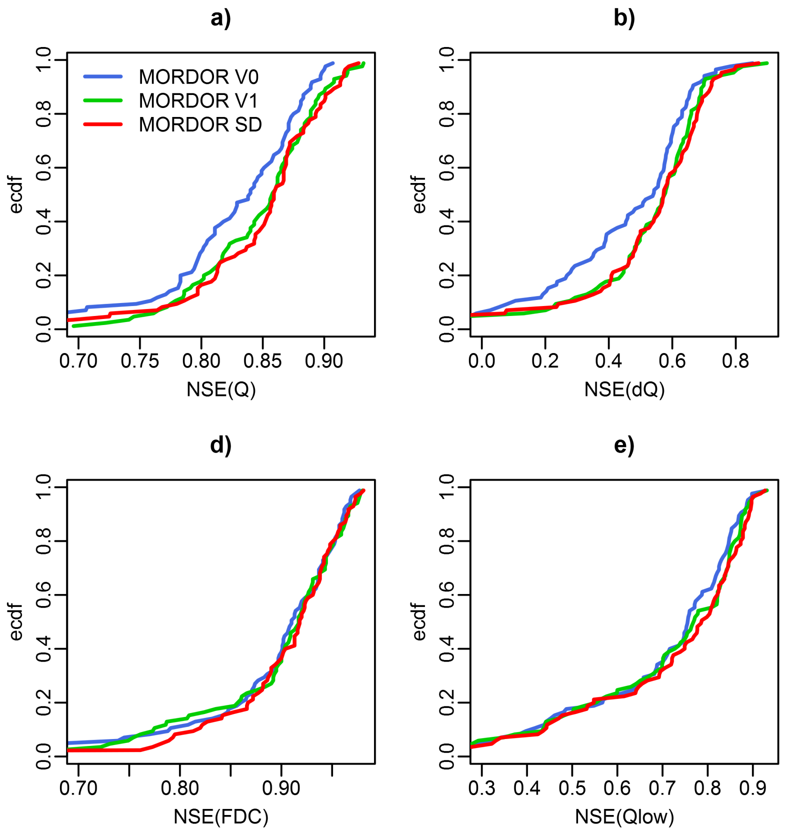

Figure 3. Performance of the three versions of the model on the validation periods, for five streamflows signatures: (a) Q; (b) dQ; (c) $Qsea$; (d) $Qlow$; (e) $Qlow$.

streamflow signatures. In contrast, taking into account orographic meteorological variability has no significant impact on model performance.

To go further, we compare the mean NSE obtained for each hydrological signature and for the three model versions. At the same time, we distinguish pluvial and nival catchments, according to the classification of Sauquet et al. (2008). The results are illustrated in Figure 4. When considering the entire dataset, we confirm previous results: MORDOR V1 and SD have very similar performance, which is significantly better than the MORDOR V0 performance, especially for Q, dQ and $Qsea$ signatures (Figure 4a). Overall, the relative improvement in performance ranges from 1% to 10%. For the pluvial catchments (Figure 4b), conclusions are the same, but overall performance is better. For nival catchments (Figure 4c), the picture clearly differs. Overall performance is lower, which underlines the high complexity of processes on these catchments. Moreover, MORDOR SD outperforms MORDOR V1 for all signatures. This improvement is especially significant for $Qlow$, Q and dQ signatures, but remains insignificant for the $Qsea$ signature. Therefore, the semi-distributed scheme clearly shows its added value for nival catchments.

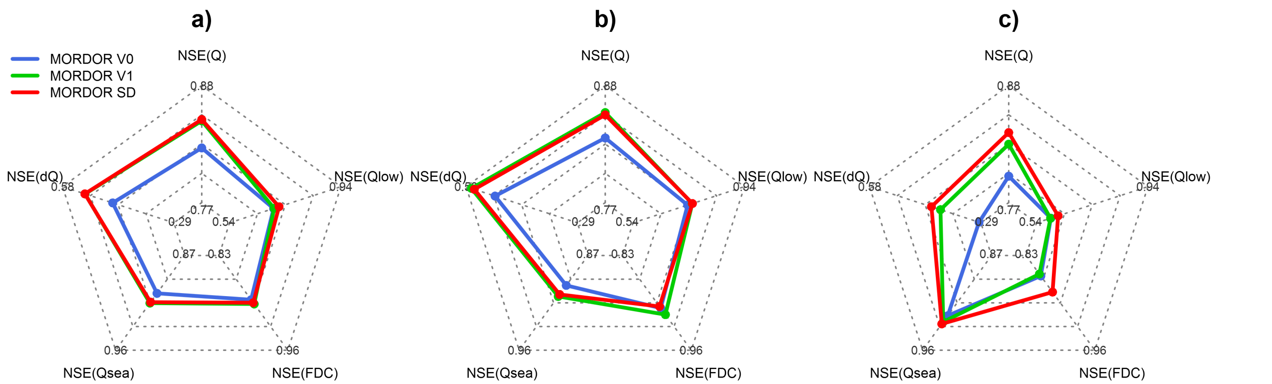

Figure 4. Mean NSE for each hydrological signature and for the three model versions: (a) for the entire catchments sample (50 catchments), (b) for the pluvial sample (35 catchments), (c) for the nival sample (15 catchments).

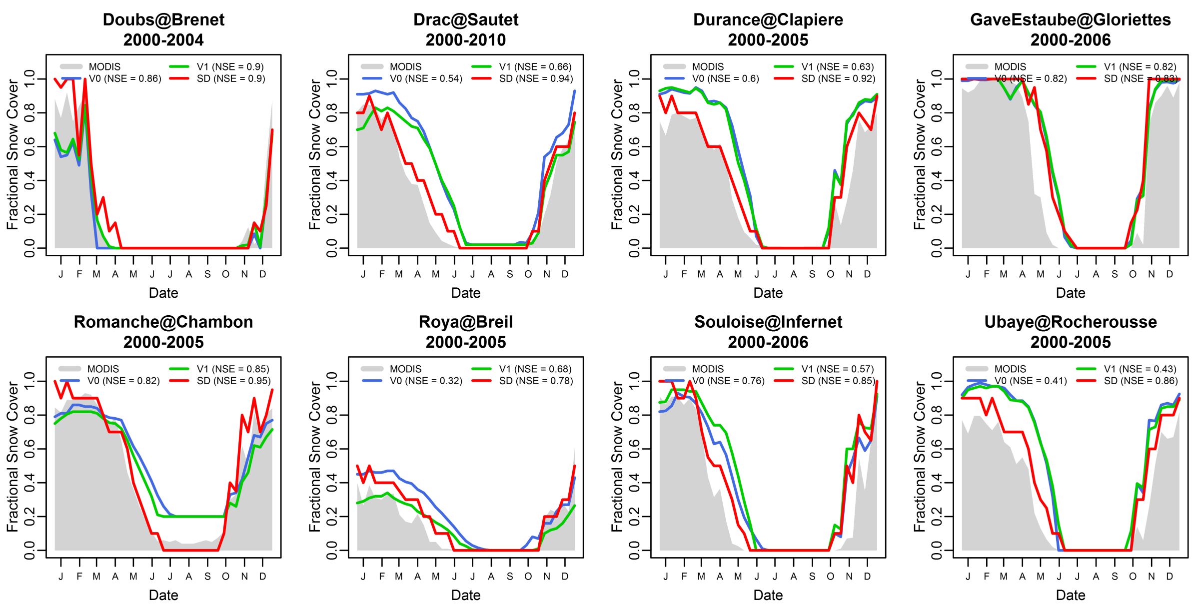

Figure 5. Fractional Snow Cover Regime on eight mountainous catchments. Comparison of MOD10 FSC product with the three model versions. For each catchment, the considered period is given. NSE values are calculated on FSC regimes.

4.2 Improvement in the representation of the snow processes

One of the objectives of this study was to improve the model representation of snow processes. Hereafter, we investigate this question using two types of data. The first one is a catchment scale average of the fractional snow cover (FSC) provided by the

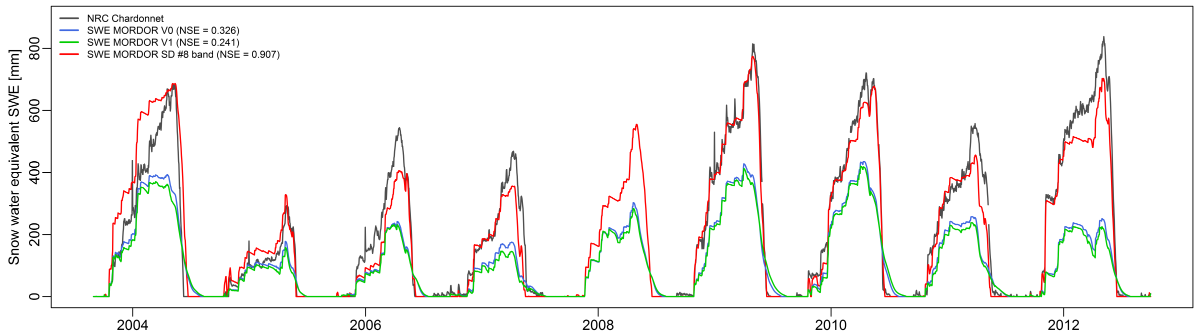

Figure 6. Observed and simulated snow water equivalent (SWE) time-series on the Durance at La Clapiere catchment. NSE values are calculated on SWE time-series.

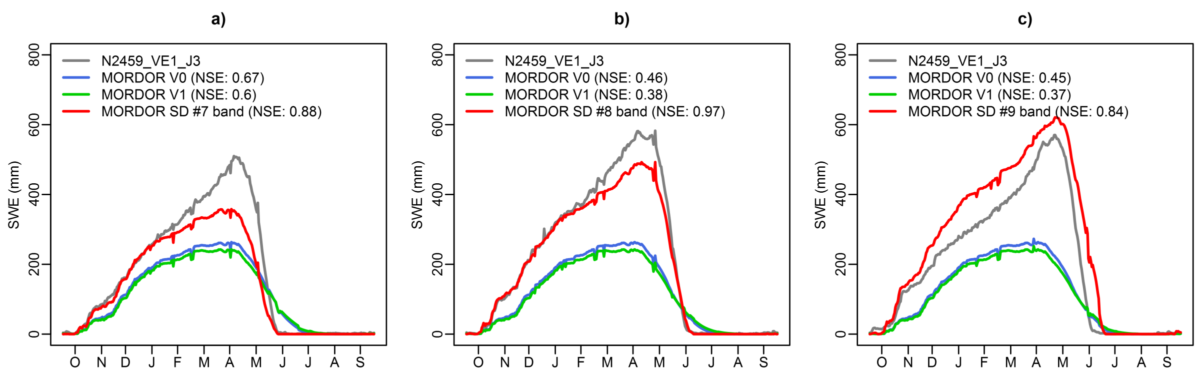

Figure 7. Observed and simulated snow water equivalent (SWE) regimes on the Durance at La Clapiere catchment, for three measurement stations: a) Izoard (2280 m a.s.l.), b) Chardonnet (2455 m a.s.l.), c) Marrous (2730 m a.s.l.). NSE values are calculated on SWE regimes.

MOD10 product, available over the 2000-2012 period. Due to uncertainties and missing data, we consider only the long-term mean daily FSC. The second one is the snow water equivalent at the local scale, derived from our NRC observation network.

Figure 5 illustrates for eight mountainous catchments the regime of the modeled and observed fractional snow cover over available periods (i.e., common periods between modeling and observations). These catchments have been selected among the nival sample (15 catchments) considering data availability. On most of these catchments, MORDOR V0 and V1 show similar behaviour, characterized by a late snow melt and an overestimation of FSC during spring and autumn. On the other hand, MORDOR SD provides a much more realistic FSC, especially during spring. Snowpack discretization within the catchment makes it possible to better represent the snow cover evolution. Finally, taking into account orographic meteorological variability significantly improves the FSC simulation, as illustrated by NSE values (see legends of Figure 5).

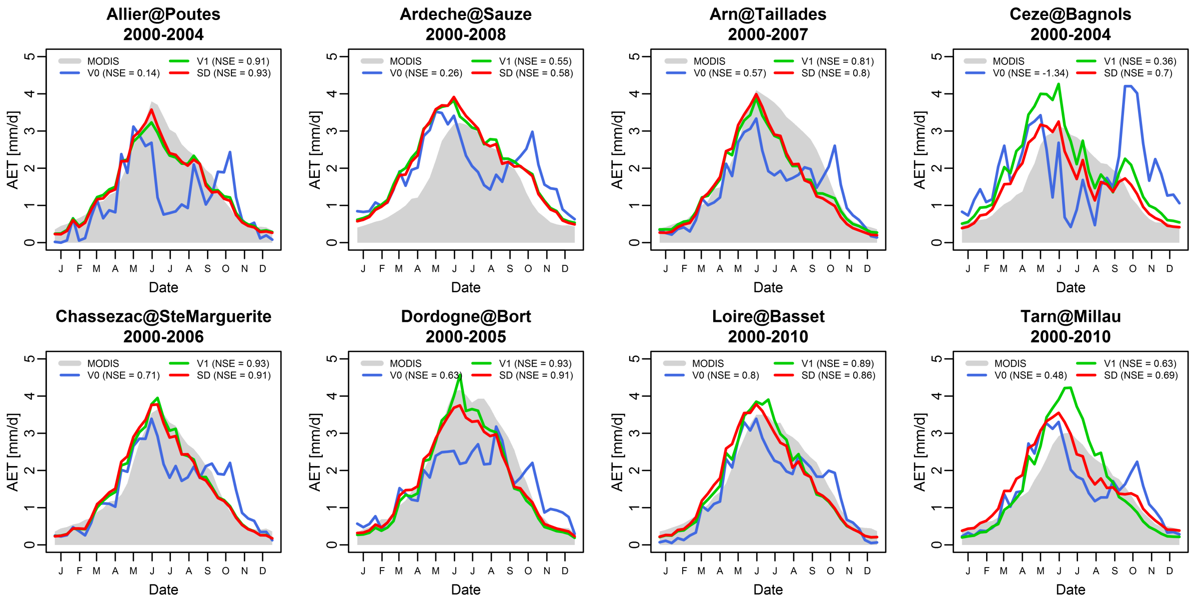

Figure 8. Actual evapotranspiration regime on eight pluvial catchments. Comparison of MOD16 AET product with the three model versions. For each catchment, the considered period is given. NSE values are calculated on AET regimes.

Figure 6 compares observed and simulated SWE time series over the Durance at La Clapiere catchment for the 2004-2012 period (observations are missing for 2008). The observations come from the Chardonnet NRC (2500 m a.s.l.). The MORDOR V0 and V1 simulations (blue and green curves) correspond to the global SWE at the catchment scale, given that they do not represent spatial variability. The MORDOR SD simulation (red curve) corresponds to elevation zone #8 situated close to the NRC altitude. First, MORDOR V0 and V1 simulations are very similar and significantly underestimate the total amount of SWE. This is a clear conceptual limitation of such global formulations which only simulate bulk values that cannot be compared to local observations. On the other hand, the semi-distributed scheme shows fairly good agreement when comparing local observations to corresponding elevation zone modeling. MORDOR SD correctly simulates the interannual variability of the maximum snowpack at this altitude, which varies from about 300 mm in 2005 to about 800 mm in 2007 and 2012. The seasonal dynamic is also very realistic, since both accumulation and melting periods are well simulated. These results are confirmed by Figure 7 for the three snow gauges located over Durance at La Clapiere catchment (see Figure 2). We compare the observed interannual SWE regime (2000-2012 period) with MORDOR V0,V1 and SD SWE. In Figures 7a, 7b and 7c, MORDOR V0 and V1 SWE,respectively, are the same and correspond to the bulk SWE at the catchment scale. In contrast, MORDOR SD SWE corresponds to #7, #8 and #9 elevation zones, respectively. Logically, the MORDOR V0 and V1 SWE underestimation increases with elevation. Instead, SWE regimes simulated by MORDOR SD are consistent with at-site observations for all elevations.

4.3 Improvement in the representation of the evapotranspiration processes

The realism of the hydrologic representation is also investigated considering the water balance, by comparing simulated ET fluxes and MOD16 satellite-derived data available over the 2000-2012 period. Due to uncertainties and missing data, we consider only the long-term mean daily ET. In addition, considering MOD16 limitations on mountainous areas, we focus on eight low altitude catchments where it may be considered as realistic. These catchments have been selected among the pluvial sample (35 catchments) considering data availability. Figure 8 shows ET regimes on the available periods (i.e., common periods between modeling and observations). Firstly, it's worth noting that PET is considered very differently for the three model structures. MORDOR V1 and SD use a PET estimated as described by (Oudin et al., 2005), which vary from 420 $mm.yr^{-1}$ to 890 $mm.yr^{-1}$ on the study catchments. On the other hand, MORDOR V0 uses an adjusted PET from temperature and model parameters which vary 220 $mm.yr^{-1}$ to 1750 $mm.yr^{-1}$. Secondly, MORDOR V1 and SD use a crop coefficient-based formulation, which is not the case for MORDOR V0. These differences have a great impact on ET regimes. Compared to the MORDOR V0 reference, ET is increased during spring and summer but decreased in autumn at the end of the growing season.

Comparison with MOD16 data suggests that this new seasonality is more realistic, as illustrated by NSE values (see legends of Figure 8). In particular it removes the unrealistic increase of ET in autumn during vegetation senescence (see for instance Allier and Ceze catchments). In this case, spatial discretization (MORDOR SD) has a second-order effect.

5 Conclusions

In this study we validated improvements in an operational hydrological model, using a multi-catchment, multi-criterion and multi-variable framework. From the historical version of the model, two alternative structures were evaluated. Within the first, the physical equations were revisited to better represent the main hydrological components, such as evapotranspiration and snow, and to reduce model parameters. The second alternative structure integrates this new formulation in an elevation zone spatialization (semi-distributed scheme).

A first evaluation focused on runoff simulation with a multi-criterion split-sample test. Five criteria were identified to focus on various streamflow signatures. For each criterion, the two alternative models perform significantly better than the initial one. On pluvial catchments, improvements are mainly due to the new physical formulation. In contrast, orographic discretization provides the main gains on nival catchments. Finally, the new semi-distributed model shows significantly better performance for runoff simulation for all catchments and for all criteria.

The second evaluation was performed on two independent hydrological variables, not used for model training: snow and evapotranspiration. The objective was to reinforce our conclusions, by performing a discharge-independent validation. The results clearly demonstrate model improvement. This semi-distributed structure simulates snow processes quite realistically. The simulation of snow cover and snow water equivalent are significantly improved. The realism of the water balance is also improved in the new model formulation. When compared with satellite proxy, the evapotranspiration dynamic is shown to be substantially improved.

This paper has therefore shown that MORDOR SD provides a very efficient tool for wide-ranging hydrological applications to hydrological simulation in pluvial and nival catchments. The performance and versatility of this new model version are very significantly improved. At the same time, its structure has been simplified, specially concerning snow processes, with fewer free parameters. Currently, further experience with MORDOR SD is being gained as it is implemented in the EDF flood-forecasting chain and in hydrological studies. An assimilation scheme is also being implemented, which integrates both discharge and snow measurement. Future work will focus upon implementation of a fully distributed version of the MORDOR SD model over large-scale catchments and in ungauged contexts.

Appendix A: MORDOR SD

This section details the MORDOR SD model structure. Figure 9 shows the wiring diagram of MORDOR SD model. It is important to underline that MORDOR V1 equations are exactly the same as MORDOR SD, differing only in that the watershed is not descritized into elevation zones.

A1 Watershed description

The MORDOR SD model is based on a succinct description of the catchment, through the following characteristics: (i) sbv the watershed area [km²]; (ii) f_{ice} relative ice area [%]; (iii) f_{lake} relative lake area [%]; (iv) $xlat$ latitude of the watershed centroid [°]; (v) $\bar{fl}$ the mean of flow length of each gridcell to the outlet [km]; and (vi) the average elevation of watershed $\bar{z}$. Furthermore, the watershed is descretized into several elevation zones. Each zone i is described by its relative area s_i [%] and its median elevation z_i [m]. Implicitly $\sum_{i=1}^{i=N_b} s_i = 1$, where i is the zone index and n_b is the total number of zones. n_b is equal to 1 in the case of MORDOR V1.

A2 Forcing

The model has as input data, for each elevation zone i and timestep t, three forcings: (i) precipitation $P_i(t)$ [mm]; (ii) air temperature $T_i(t)$ [°C]; and (iii) potential evapotranspiration $PET_i(t)$ [mm]. Often in the operational context only the areal precipitation $P(t)$ and temperature $T(t)$ are available. In this case, the forcing data for each zone are computed through two orographic gradients:

$$P_i(t) = P(t) \cdot (1 + \frac{gpz}{1000}) \cdot (z_i - \bar{z}) \quad \text{(A1)}$$

$$T_i(t) = T(t) + \frac{gtz}{100} \cdot (z_i - \bar{z}) \quad \text{(A2)}$$

where gpz is the precipitation gradient [%/1000 m] and gtz is the temperature gradient [°/100 m]. In this case, the $PET_i(t)$ could be computed with several formulas driven by $T_i(t)$, for instance following the formula proposed by Oudin et al. (2005). These equations are not used in MORDOR V1.

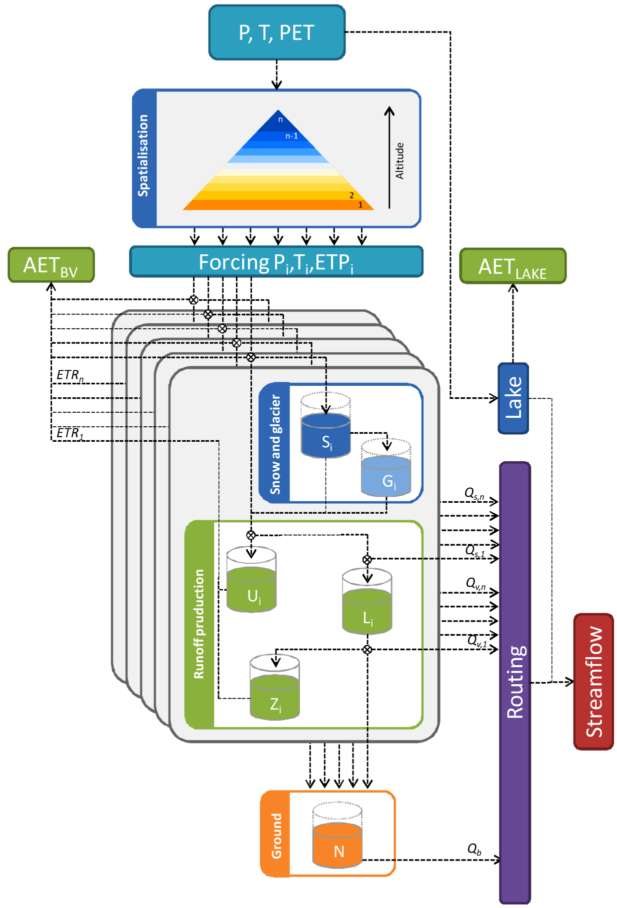

Figure 9. Overview of MORDOR SD model components and fluxes.

A3 Water balance

From the potential evapotranspiration $PET_i(t)$, a maximum evapotranspiration $MET_i(t)$ is computed using a crop coefficient Kc, such as:

$$MET_i(t) = c_{etp} \cdot Kc(t) \cdot PET_i(t) \tag{A3}$$

where c_{etp} [-] is a correction factor of the total amount of PET. In its classical form, the Kc coefficient varies during the growing season and is defined for any crop using look-up tables (Allen et al., 1998). However, in an operational and meso-scale context, a watershed-effective Kc must be defined, in order to accommodate various hydrological contexts and to efficiently supply the water balance. In the model, the Kc formulation is:

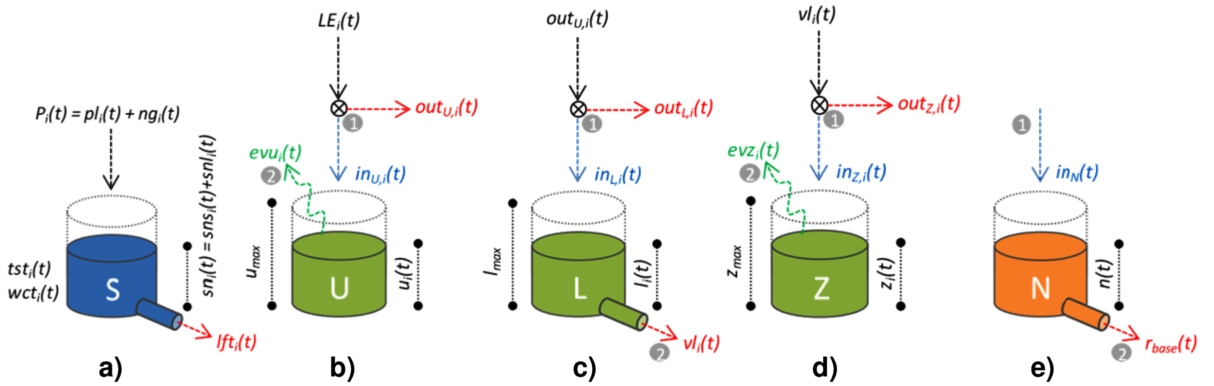

Figure 10. Schematic representation of MORDOR SD storage.

$$K_c^i(t) = K_{min} + (1 - K_{min}) \cdot \frac{(Rpot_i(t) - \min\{Rpot(t)\})}{(\max\{Rpot(t)\} - \min\{Rpot(t)\})} \tag{A4}$$

$$K_c^i(t) = \frac{K_c^i(t)}{\bar{K}_c^i} \tag{A5}$$

with K_{min} the minimum seasonal crop coefficient value and $Rpot$ $[W \cdot m^{-2} \cdot day{-}1]$ the potential solar radiation. From the MET, the model calculates the actual evapotranspiration (AET) from three components: (i) surface interception $ev0_i(t)$ according to the following formula: $ev0_i(t) = min(MET_i(t), P_i(t))$; (ii) evapotranspiration from the root soil $evu_i(t)$, see A5.1; and (iii) evapotranspiration from the capillarity water storage $evz_i(t)$, see A5.3.

A4 Snow module

The aim of storage S is to model the snow pack. Figure 10a shows the I/O and the state variables of this storage.

A4.1 Snow accumulation

For each elevation zone i and timestep t, the precipitation $P_i(t)$ is divided into two components: (i) the liquid part $pl_i(t)$, i.e., rain, and (ii) the solid part $ng_i(t)$, i.e., snow. Then the inputs of the storage S are:

$$pl_i(t) = fliq(t) \cdot P_i(t) \tag{A6}$$

$$ng_i(t) = (1 - fliq(t)) \cdot P_i(t) \tag{A7}$$

where $fliq(t)$ is the liquid ratio of precipitation founded on the classical parametric S-shaped curve:

$$fliq_i(t) = 1 - [1 + exp(\frac{10}{\delta T} \cdot ((T_i(t) + efp) - t_{50}))]^{-1}] \tag{A8}$$

where δT is the thermic range (set to 4 °C), t_{50} is the threshold temperature between the solid and liquid phases (set to 1 °C) and efp [°C] is an additive correction parameter, by default set equal to zero.

A4.2 Snow melt

For each elevation zone i and each timestep t, the snow pack is summarized by two state variables: the bulk temperature $tst_i(t)$ and the water content $wct_i(t)$. The snow pack temperature is computed using an exponential smoothing function as follows:

$$tst_i(t) = \min\{lts \cdot tst_i(t-1) + (1-lts) \cdot (T_i(t) + efp), 0\} \tag{A9}$$

where lts [-] is the smoothing parameter between the antecedent snow pack temperature and the actual modified air temperature. The melt runoff $lft_i(t)$ is composed of two parts: the surface melt $lft_i(t)$ and the ground melt gm. The latter is considered constant in time and space. The surface melt changes according to the elevation zone i and the timestep t as follows:

$$lfs_i(t) = K_f(t) \cdot (T_i(t) + eft + tst_i(t)) \tag{A10}$$

where K_f is the melting coefficient and eft [°C] is the additive correction parameter, by default set to zero. The melting coefficient $K_f(t)$ is computed via this equation:

$$K_f(t) = k_f + (k_{fp} \cdot \frac{R_{pot}(t)}{\bar{R_{pot}}}) \tag{A11}$$

where k_f is the constant part and k_{fp} the variable part of the melting coefficient, indexed on potential solar radiation. For a given elevation zone i and the timestep t, the output of the snow model is the runoff $le_i(t)$ equal to the sum of the rainfall $p_i(t)$, the surface melt $lfs_i(t)$ and the ground melt gm.

A5 Runoff production

A5.1 Surface storage U

The storage U is intended to represent the water absorption capacity of the root zone. As shown in Figure 10b, the I/O of storage U follows these equations:

$$in_{U,i}(t) = (u_{max} - u_i(t-1)) \cdot (1 - \exp(-\frac{le_i(t)}{u_{max}})) \tag{A12}$$

$$out_{U,i} = le_i(t) - in_{U,i}(t) \tag{A13}$$

$$evu_i(t) = (MET_i(t) - ev0_i(t)) \cdot \frac{u_i(t)}{u_{max}} \tag{A14}$$

where $u_i(t)$ is the water content of storage U for the elevation zone i at the timestep t and u_{max} [mm] is the maximum capacity of the storage, assumed constant for all zones. This parameter is assumed to be the same for all zones.

A5.2 Hillslope storage L

The storage L is intended to represent the hillslope zone. As shown in Figure 10c, the I/O of storage L follows these equations:

$$in_{L,i}(t) = out_{U,i}(t) \cdot [1 - (\frac{l_i(t-1)}{l_{max}})^2] \quad (A15)$$

$$out_{L,i} = out_{U,i}(t) - in_{L,i}(t) \quad (A16)$$

$$v_{L,i}(t) = k_L \cdot l_i(t)^{evl} = \frac{1}{evl \cdot l_{max}^{evl-1}} \cdot l_i(t)^{evl} \quad (A17)$$

where $l_i(t)$ is the water content of storage L for the elevation zone i at the timestep t and l_{max} [mm] is the maximum capacity of the storage, assumed constant for all zones. The parameter evl [-] is the outflow exponent. Then the surface runoff, $r_{surf,i}(t)$, provided by the elevation zone i is computed according to:

$$r_{surf,i}(t) = out_{L,i}(t) - max(0, in_{L,i}(t) - l_{max}) \quad (A18)$$

A5.3 Capillarity storage Z

The storage Z is intended to represent the capillarity of the hillslope zone. As shown in Figure 10d, the I/O of storage Z follows these equations:

$$in_{Z,i}(t) = v_{L,i}(t) \cdot [1 - (\frac{z_i(t-1)}{z_{max}})] \quad (A19)$$

$$out_{Z,i} = v_{L,i}(t) - in_{Z,i}(t) \quad (A20)$$

$$evz_i = (MET_i(t) - ev0_i(t) - evu_i(t)) \cdot \frac{z_i(t)}{z_{max}} \quad (A21)$$

where $z_i(t)$ is the water content of storage Z for the elevation zone i at the timestep t and z_{max} [mm] is the maximum capacity of the storage, assumed constant for all zones. Then the subsurface runoff, $r_{vers,i}(t)$, provided by the elevation zone i, is computed according to:

$$r_{vers,i}(t) = k_r \cdot out_{Z,i}(t) \quad (A22)$$

where k_r [-] is the runoff coefficient, ranging from 0 to 1.

A5.4 Ground storage N

The deep storage N determines the baseflow runoff. As shown in Figure 10e, the I/O of storage N follows these equations:

$$in_N(t) = \sum_{i=1}^{N_b} ((1-k_r) \cdot out_{Z,i}(t)) \cdot s_i \quad (A23)$$

$$r_{base}(t) = k_N \cdot n(t)^{evn} \quad (A24)$$

where the parameter k_N $[mm \cdot hour^{-1}]$ is the outflow coefficient and and the parameter evn [-] the outflow exponent.

A6 Routing function

The model identifies three flux components: (i) surface runoff r_{surf}; (ii) subsurface exfiltration r_{vers}; (iii) base flow r_{base}. The global streamflow $rt(t)$ is the sum of these three components, as follows:

$$rt(t) = (\sum_{i=1}^{N_b} r_{surf,i}(t)) \cdot s_i) + (\sum_{i=1}^{N_b} r_{vers,i}(t)) \cdot s_i) + r_{base}(t) \tag{A25}$$

The routing function used to transfer the global streamflow to the outlet is based on the diffusive wave equation (Hayami, 1951):

$$f(t, cel, dif) = \frac{\bar{fl}}{2\sqrt{\pi dif}} \cdot t^{-\frac{3}{2}} \cdot e^{-\frac{(\bar{fl} - cel \cdot t)^2}{4 \cdot dif \cdot t}} \tag{A26}$$

where t is the timestep, cel is the celerity of the wave $[km/h]$ and dif is the diffusion of the wave in $[km^2/h]$.

A7 MORDOR SD parameters overview

Table 2 summarizes the 19 free parameters of MORDOR SD model.

A8 Technical details

The MORDOR SD model is written in FORTRAN 90. The model runs at different temporal resolution. The duration of a simple model simulation (i.e., model run and evaluation criteria computation) is approximately 1 sec and depends on the time step and on the length of time series. For instance a daily simulation over 50 years takes less than 1 sec and an hourly simulation over 10 years takes approximately 2 sec. Concerning the calibration process (approximately 40,000 model runs), the algorithm takes approximately 10 min for a daily time step over 50 years and approximately 45 min for an hourly time step over 10 years. The post-processing and graphical tools are developed in R language.

Table 2. MORDOR SD free parameters, units, range and description.

Parameter	Units	Prior range	Description
cetp	-	[0.8,1.2]	PET correction factor
gtz	$^{\circ}C/100m$	[-0.8,-0.4]	Air temperature gradient
gpz	$\%/1000m$	[0.1,0.7]	Precipitation gradient
kmin	-	[0.1,1.5	Minimum seasonal crop coefficient
umax	mm	[30,400]	Maximum capacity of the root zone
zmax	mm	[30,400]	Maximum capacity of the capillarity storage
lmax	mm	[30,400]	Maximum capacity of the hillslope zone
evl	-	[1.5,4]	Outflow exponent of storage L
kr	-	[0.1,0.9]	Runoff coefficient
kn	$mm \cdot hour^{-1}$	[10^{-10},10^{-1}]	Outflow coefficient of storage N
evn	-	[1,4]	Outflow exponent of storage N
kf	$mm \cdot ^{\circ}C^{-1} \cdot day^{-1}$	[0,5]	Constant part of melting coefficient
kfp	$mm \cdot ^{\circ}C^{-1} \cdot day^{-1}$	[0,5]	Variable part of melting coefficient
efp	$^{\circ}C$	[-3,3]	Additive correction of temperature for rain/snow partitioning
eft	$^{\circ}C$	[-3,3]	Additive correction of snow pack temperature
lts	-	[0.75,0.99]	Smoothing parameter of snow pack temperature
gm	$mm \cdot day^{-1}$	[0.4,0.8]	Ground melt
cel	$km \cdot hour^{-1}$	[0.1,10]	Wave celerity
dif	$km^2 \cdot hour^{-1}$	[0.1,50]	Wave diffusion

References

Allen, R. G., Pereira, L. S., Raes, D., Smith, M., et al.: Crop evapotranspiration-Guidelines for computing crop water requirements-FAO Irrigation and drainage paper 56, FAO, Rome, 300, D05 109, 1998.

Beldring, S.: Multi-criteria validation of a precipitation–runoff model, Journal of Hydrology, 257, 189–211, 2002.

Bergstroem, S.: The development of a snow routine for the HBV-2 model, Hydrology Research, 6, 73–92, 1975.

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.

Breuer, L., Huisman, J., Willems, P., Bormann, H., Bronstert, A., Croke, B., Frede, H.-G., Gräff, T., Hubrechts, L., Jakeman, A., et al.: Assessing the impact of land use change on hydrology by ensemble modeling (LUCHEM). I: Model intercomparison with current land use, Advances in Water Resources, 32, 129–146, 2009.

Chahinian, N., Andreassian, V., Duan, Q., Fortin, V., Gupta, H., Hogue, T., Mathevet, T., Montanari, A., Moretti, G., Moussa, R., et al.: Compilation of the MOPEX 2004 results, IAHS publication, 307, 313, 2006.

DHI, M.: A Modelling System for Rivers and Channels, Reference Manual. DHI Software, 2009.

Duan, Q., Schaake, J., Andreassian, V., Franks, S., Goteti, G., Gupta, H. V., Gusev, Y., Habets, F., Hall, A., Hay, L., et al.: Model Parameter Estimation Experiment (MOPEX): An overview of science strategy and major results from the second and third workshops, Journal of Hydrology, 320, 3–17, 2006.

Euser, T., Winsemius, H., Hrachowitz, M., Fenicia, F., Uhlenbrook, S., and Savenije, H.: A framework to assess the realism of model structures using hydrological signatures, Hydrology and Earth System Sciences, 17, 1893–1912, 2013.

Garavaglia, F., Gailhard, J., Paquet, E., Lang, M., Garçon, R., and Bernardara, P.: Introducing a rainfall compound distribution model based on weather patterns sub-sampling, Hydrology and Earth System Sciences Discussions, 14, p–951, 2010.

Garçon, R.: Prévision opérationnelle des apports de la Durance à Serre-Ponçon à l'aide du modèle MORDOR. Bilan de l'année 1994-1995, La Houille Blanche, pp. 71–76, 1996.

Gharari, S., Hrachowitz, M., Fenicia, F., and Savenije, H.: An approach to identify time consistent model parameters: sub-period calibration, Hydrology and Earth System Sciences, 17, 149–161, 2013.

Gottardi, F., Obled, C., Gailhard, J., and Paquet, E.: Statistical reanalysis of precipitation fields based on ground network data and weather patterns: Application over French mountains, Journal of Hydrology, 432, 154–167, 2012.

Gupta, H. V., Kling, H., Yilmaz, K. K., and Martinez, G. F.: Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling, Journal of Hydrology, 377, 80–91, 2009.

Gupta, H. V., Perrin, C., Blöschl, G., Montanari, A., Kumar, R., Clark, M., and Andréassian, V.: Large-sample hydrology: a need to balance depth with breadth, Hydrology and Earth System Sciences, 18, 463, 2014.

Hall, D. K., Riggs, G. A., Salomonson, V. V., DiGirolamo, N. E., and Bayr, K. J.: MODIS snow-cover products, Remote sensing of Environment, 83, 181–194, 2002.

Hayami, S.: On the propagation of flood waves, 1951.

Henderson-Sellers, A., Yang, Z., and Dickinson, R.: The project for intercomparison of land-surface parameterization schemes, Bulletin of the American Meteorological Society, 74, 1335–1349, 1993.

Holländer, H. M., Blume, T., Bormann, H., Buytaert, W., Chirico, G., Exbrayat, J.-F., Gustafsson, D., Hölzel, H., Kraft, P., Stamm, C., et al.: Comparative predictions of discharge from an artificial catchment (Chicken Creek) using sparse data, Hydrology and Earth System Sciences, 13, 2069–2094, 2009.

Hrachowitz, M., Fovet, O., Ruiz, L., Euser, T., Gharari, S., Nijzink, R., Freer, J., Savenije, H., and Gascuel-Odoux, C.: Process consistency in models: The importance of system signatures, expert knowledge, and process complexity, Water resources research, 50, 7445–7469, 2014.

Hu, G., Jia, L., and Menenti, M.: Comparison of MOD16 and LSA-SAF MSG evapotranspiration products over Europe for 2011, Remote Sensing of Environment, 156, 510–526, 2015.

Kavetski, D. and Fenicia, F.: Elements of a flexible approach for conceptual hydrological modeling: 2. Application and experimental insights, Water Resources Research, 47, 2011.

Kirchner, J. W.: Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology, Water Resources Research, 42, 2006.

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

Kodama, M., Nakai, K., Kawasaki, S., and Wada, M.: An application of cosmic-ray neutron measurements to the determination of the snow-water equivalent, Journal of Hydrology, 41, 85–92, 1979.

Mathevet, T.: Quels modeles pluie-debit globaux au pas de temps horaire? Développements empiriques et intercomparaison de modeles sur un large échantillon de bassins versants, Ph.D. thesis, Ph. D. thesis, ENGREF, 463 pp, 2005.

Mathevet, T. and Garçon, R.: Tall tales from the hydrological crypt: are models monsters?, Hydrological Sciences Journal–Journal des Sciences Hydrologiques, 55, 857–871, 2010.

Micovic, Z. and Quick, M.: A rainfall and snowmelt runoff modelling approach to flow estimation at ungauged sites in British Columbia, Journal of Hydrology, 226, 101–120, 1999.

Miralles, D., Jiménez, C., Jung, M., Michel, D., Ershadi, A., Mc-Cabe, M., Hirschi, M., Martens, B., Dolman, A., Fisher, J., et al.: The WACMOS-ET project–Part 2: Evaluation of global terrestrial evaporation data sets, Hydrol. Earth Syst. Sci. Discuss, 12, 10 651–10 700, 2015.

Motovilov, Y. G., Gottschalk, L., Engeland, K., and Rodhe, A.: Validation of a distributed hydrological model against spatial observations, Agricultural and Forest Meteorology, 98, 257–277, 1999.

Mu, Q., Zhao, M., and Running, S. W.: Improvements to a MODIS global terrestrial evapotranspiration algorithm, Remote Sensing of Environment, 115, 1781–1800, 2011.

Nijzink, R. C. and Savenije, H. H.: The importance of topography-controlled sub-grid process heterogeneity and semi-quantitative prior constraints in distributed hydrological models, Hydrology and Earth System Sciences, 20, 1151, 2016.

Orth, R., Staudinger, M., Seneviratne, S. I., Seibert, J., and Zappa, M.: Does model performance improve with complexity? A case study with three hydrological models, Journal of Hydrology, 523, 147–159, 2015.

Oudin, L., Hervieu, F., Michel, C., Perrin, C., Andréassian, V., Anctil, F., and Loumagne, C.: Which potential evapotranspiration input for a lumped rainfall–runoff model?: Part 2—Towards a simple and efficient potential evapotranspiration model for rainfall–runoff modelling, Journal of hydrology, 303, 290–306, 2005.

Paquet, E. and Laval, M.-T.: Retour d'expérience et perspectives d'exploitation des Nivomètres à Rayonnement Cosmique d'EDF, La Houille Blanche, pp. 113–119, 2006.

Paquet, E., Garavaglia, F., Garçon, R., and Gailhard, J.: The SCHADEX method: A semi-continuous rainfall–runoff simulation for extreme flood estimation, Journal of Hydrology, 495, 23–37, 2013.

Parajka, J. and Blöschl, G.: Validation of MODIS snow cover images over Austria, Hydrology and Earth System Sciences Discussions, 3, 1569–1601, 2006.

Perrin, C., Michel, C., and Andréassian, V.: Does a large number of parameters enhance model performance? Comparative assessment of common catchment model structures on 429 catchments, Journal of Hydrology, 242, 275–301, 2001.

Perrin, C., Michel, C., and Andréassian, V.: Improvement of a parsimonious model for streamflow simulation, Journal of Hydrology, 279, 275–289, 2003.

Reed, S., Koren, V., Smith, M., Zhang, Z., Moreda, F., Seo, D.-J., and Participants, D.: Overall distributed model intercomparison project results, Journal of Hydrology, 298, 27–60, 2004.

Rodell, M. and Houser, P.: Updating a land surface model with MODIS-derived snow cover, Journal of Hydrometeorology, 5, 1064–1075, 2004.

Sauquet, E., Gottschalk, L., and Krasovskaia, I.: Estimating mean monthly runoff at ungauged locations: an application to France, Hydrology Research, 39, 403–423, 2008.

Smith, M. B., Koren, V., Reed, S., Zhang, Z., Zhang, Y., Moreda, F., Cui, Z., Mizukami, N., Anderson, E. A., and Cosgrove, B. A.: The distributed model intercomparison project–Phase 2: Motivation and design of the Oklahoma experiments, Journal of Hydrology, 418, 3–16, 2012.

Staudinger, M., Stahl, K., Seibert, J., Clark, M., and Tallaksen, L.: Comparison of hydrological model structures based on recession and low flow simulations, Hydrology and Earth System Sciences, 15, 3447–3459, 2011.

Thirel, G., Salamon, P., Burek, P., and Kalas, M.: Assimilation of MODIS snow cover area data in a distributed hydrological model using the particle filter, Remote Sensing, 5, 5825–5850, 2013.

Trambauer, P., Dutra, E., Maskey, S., Werner, M., Pappenberger, F., Van Beek, L., and Uhlenbrook, S.: Comparison of different evaporation estimates over the African continent, Hydrology and Earth System Sciences, 18 (1), 2014, 2014.

Valéry, A., Andréassian, V., and Perrin, C.: 'As simple as possible but not simpler': What is useful in a temperature-based snow-accounting routine? Part 2–Sensitivity analysis of the Cemaneige snow accounting routine on 380 catchments, Journal of Hydrology, 517, 1176–1187, 2014a.

Valéry, A., Andréassian, V., and Perrin, C.: 'As simple as possible but not simpler': What is useful in a temperature-based snow-accounting routine? Part 2–Sensitivity analysis of the Cemaneige snow accounting routine on 380 catchments, Journal of Hydrology, 517, 1176–1187, 2014b.

Vansteenkiste, T., Tavakoli, M., Van Steenbergen, N., De Smedt, F., Batelaan, O., Pereira, F., and Willems, P.: Intercomparison of five lumped and distributed models for catchment runoff and extreme flow simulation, Journal of Hydrology, 511, 335–349, 2014.

Wang, Q.: The genetic algorithm and its application to calibrating conceptual rainfall-runoff models, Water resources research, 27, 2467–2471, 1991.

Westerberg, I. and McMillan, H.: Uncertainty in hydrological signatures, Hydrology and Earth System Sciences, 19, 3951–3968, 2015.

Zanotti, F., Endrizzi, S., Bertoldi, G., and Rigon, R.: The GEOTOP snow module, Hydrological Processes, 18, 3667–3679, 2004.