The authors express their deep gratitude to the referee for a thorough analysis of the article and a significant number of valuable comments. Criticism of the English language of the manuscript is fully accepted, the authors will make efforts to correct the text. The following are responses to individual comments.

This manuscript details a study that compares runoff decomposition as estimated by end-member mixing analysis on one hand and hydrological models on the other hand. This issue is clearly of great interest for both hydrological processes understanding and model development. At this stage, the paper has several shortcomings in terms of methods and data, which prevents a full understanding of the paper. My suggestion would be to be less ambitious, e.g. in the number of hydrological models used but more exhaustive in the details given throughout the manuscript. Additionally, it should be noted that the English level is pretty poor, I am not a native English speaker but I recommend that the authors proofread their revised manuscript before submission.

The main idea of the article was precisely that several widely used runoff models are compared, which are parameterized by conventional, standard methods, completely independently of one to another and something else. Initially the authors decided not to pay too much attention to the standard descriptions, it to be done. Then, from the resulting simulated series, data on runoff components are extracted, which are direct depending on the model used. These data are compared with runoff components obtained from EMMA, on the basis of which the authors try to draw conclusions about the greater or lesser adequacy of different models. The question that the article tries to answer is the following: is such an approach promising in principle? Thus, the comparison of just a set of models seems to be the key for the author's approach.

Major comments

Lack of details

Throughout the paper, there is a lack of details. This affects both the data/method section and the results/discussion section at a level that precludes the reader to provide clear guidelines for further improvements. The required additional information is listed in the minor comments hereafter. The two other major comments are related to methodological issues.

The comment is accepted, the text will be improved.

Hydrological model uncertainties and how the methodology of the paper reduces it

As explained in the introduction (l.31-39), hydrograph decomposition may be a powerful tool to reduce equifinality. In this sense, the present study shows relatively similar runoff simulations but with different flow components from hydrological models, but with different flow components. Unfortunately, the authors did not take this opportunity seriously, they used a single optimal parameter set for each model and did not discuss the impact of this choice, nor the way the parameter set is optimized. Consequently, it is pretty hard to conclude the relative weights of structural and parameter uncertainties in explaining the results.

The possibilities of the authors' analysis are limited by the volume of available information. For each case, only three years were used when comparing data for both runoff modeling and EMMA analysis. Although there is a little more data only for calibration, it is still not enough for various fine measurement methods, comparison of parameter sets, etc. e. This shortcoming the authors try to compensate by using several models and taking into account the composition of the runoff.

Short record periods

Only three years are available for model simulation (what about the warm-up year?). It is quite short and consequently, no validation was performed by the authors. Modeling results are presented only for calibration, which is problematic when dealing with parameter/structural uncertainties. Also, as low-flow components are extracted from hydrological models simulations, the authors should verify cautionly model initialization.

The answer to the question is partially given above. An analysis of the literature shows that data sets that allow continuous daily runoff decomposition over several years are extremely rare. Thus, the noted lack of data is an objective circumstance and should be overcome in the course of further work.

Four years including warm-up for each catchments were used. One year for warm-up definitely should be enough for properly initialize small catchment models.

Minor (but still important) comments

l.77-78: not clear what is the time step of heavy rainfall and how extreme are these events.

In the description of the objects of study, we indicate:
l.73: "The annual average precipitation amount is 700-800 mm"
l.78: "The range of maximal daily heavy rains is 100–200 mm."
Thus, by comparing the average annual precipitation and for an individual rain, one can draw a conclusion about the extremeness of such events.

l.79: not clear how averaging is performed, spatial or temporal?

It means sum (temporal) of precipitation for rain, measured by rain-gauge.

Please add a table with both catchments characteristics (mean annual rainfall, temperature, runoff, land use lithology, topography, etc.). The differences in runoff yields for these two neighbor catchments are huge and I cannot figure out if it is due to lithological differences or specificity of the (short) record periods with extreme events.

This notes to be taken into account when editing.
We present the table 1 following with information about catchments.
The differences are rather due to the geological structure than a short series of observations

Table 1. Catchment's characteristics.

Characteristics	Elovy	Medvezhy
Area, km2	3.5	7.6
Avg. Height, m	722	704
Max Height, m	962	869
Avg. slope, %	13.5	13.8
Max. slope, %	28.7	31.5
Avg. Precipitation, mm/day *	2.13	2.35
Avg. Temperature, C *	3	3.23
Avg. Discharge, mm/day *	0.65	0.75
Land-use	Coniferous-broadleved and coniferous forest	Coniferous-broadleved forest
Lihology	Cretaceous volcanites (tuffs) and sub-volcanic acid and intermediate rocks (granites, ryolite, porphyrites and diorites)	Jurassic metamorphic basic rocks (gabbroids etc.)

*daily values, assessment period: 01.01.11-31.12.14 for Elovy creek and 01.01.14-31.12.17 for Medvezhy creek

Figure 1. Where the WMO station is located?

This will be taken into account when editing. Since the weather station is located 30 kilometers away, we add the coordinates: WMO 31939 (Chuguevka) weather station located in 35 km to the NW from the observation sites (E 133°53'48" N 44°11'59").

l.132: It is not clear how the end members are identified, what is the "independent information"?

In this case, it's a lapse. This meant (but wrongly expressed) the data on potential sources that were not included in the series of river water data. It will be edited.

Please add a table with the characteristics of the three hydrological models (with e.g. basis of the snow components, number of free parameters, spatial and temporal discretization, etc.). The information given for each model is not homogenous. Nothing is said on parameter estimation, which is in my opinion a key issue (see major comment #2).

The authors will make an effort to present homogeneous information on the three applied hydrological models (Table 2).

Table 2. Hydrological models characteristics

Characteristics / Model	ECOMAG	SWAT	HBV
Spatial discretization	HRU	HRU	Lumped
Temporal discretization	Daily	Daily	Daily
Number of calibrated parameters	9	12	10
Snow component	Degree-day	Degree-day	Degree-day
Evaporation	Dalton method	Penman-Monteith	Penman-Monteith
Surface flow	Kinematic wave	Kinematic wave	Storage based
Soil flow	Darcy's law	kinematic storage model	Storage based
Groundwater flow	Darcy's law	Storage based	Storage based
Routing	Kinematic wave	Variable travel time	Triangular weighted

l.238-249: Are the results are shown in calibration (and by the way, how the calibration is performed)?? Please modify Figure 5 so that the reader can see the whole record period simulation results.

The remarks above relate to the design of materials and the detail of the presentation of the article. In general, comments are accepted, the text of the article will be improved taking into account these shortcomings.

Fig. 5 shown results of calibration. Calibration was performed manually for all models. Figure 5 was modified to represent all simulation period (without warm-up year).

Table 4: it appears that the models present quite different flow decompositions. Could this be due to the fact that the a priori three-component is wrong because too detailed for such small catchments? Since many flow decompositions only concern two flow components ("baseflow" and "surface flow"), did the authors challenge their prior 3-components hypothesis?

The hypothesis of three main runoff components (conditionally "direct flow", "soil flow" and "ground flow") is generally accepted and widely confirmed in tracer studies based on EMMA. The text of the article indicates that the EMMA analysis for the studied basins quite clearly showed three runoff components in one case and four in the other. As for simulation models, the number of runoff components in them is determined by the structure of the model. Only 3-component runoff models were specially selected for comparability of simulation results with EMMA decomposition. The variety of runoff composition indicated by the reviewer, obtained in various models, is precisely the key problem considered in the article. For larger basins, this diversity is no less than for the smallest ones. The indicated problems are also strongly conditioned by the lack of unification of theoretical ideas about the runoff and the corresponding terminology, which is discussed in sufficient detail in the text of the article. This lack cannot be compensated within the scope of this article and must be taken into account in the evaluation of the results.

l.359-360: These perspectives are quite fuzzy. Please provide a real discussion section in the paper. There is a lot to say on both methodological limitations and further analysis of the results obtained.

The Discussion section will be checked and edited.